DOCUMENT RESUME

ED 280 872 TM 870 202

TITLE	Plans and Activities for 1990 Decennial Census. Part 2: Hearings before the Subcommittee on Census and Population of the Committee on Post Office and Civil Service. House of Representatives, Ninety-Ninth Congress, Second Session (May 1, 15, 1986).
INSTITUTION	Congress of the U. S., Washington, D. C. House Committee on Post Office and Civil Service.
PUB DATE	May 86
NOTE	214p.; Serial No. 99-48. Some sections contain small print.
AVAILABLE FROM	Superintendent of Documents, Congressional Sales Office, U.S. Government Printing Office, Washington, DC 20402.
PUB TYPE	Legal/Legislative/Regulatory Materials (090)
EDRS PRICE	MF01/PC09 Plus Postage.
DESCRIPTORS	*Census Figures; *Data Collection; *Data Processing; *Demography; Federal Programs; Hearings; Measurement Objectives; *National Surveys; Population Trends; Questionnaires; *Research Design; Test Construction
IDENTIFIERS	Bureau of the Census; *Census 1990; Congress 99th

ABSTRACT

Hearings on the 1990 Decennial Census were held on May 1, 1986, and May 15, 1986. The May 1 session focused on data processing procedures. Speakers included John G. Keane, Daniel G. Horvitz, William Eddy, Judith S. Rowe, Benjamin F. King, and Stephen E. Fienberg. Topics included automation of address files and questionnaire check-in; dissemination of data on a variety of media including microfiche and CD-ROM; cost effectiveness of proposed procedures; linking of household records; and use of microcomputers for data processing. Two areas were mentioned in which the Subcommittee could best assist the Bureau of the Census: the procurement of computer equipment and the oversight of plans for adjustment of censal counts. The second hearing, held on May 15, concerned the census questionnaire and automation. Topics included the design of a shorter census questionnaire form; data conversion methods such as Film Optical Sensing Device for Input to Computer (FOSDIC) and optical mark recognition (OMR); cost effectiveness; and the National Content Test. Speakers included Susan Miskura, Gene Dodaro, and Gail Franke. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED280872

PLANS AND ACTIVITIES FOR 1990 DECENNIAL CENSUS

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to: TM

In our judgment, this document is also of interest to the Clearinghouses noted to the right. Indexing should reflect their special points of view. SO

PART 2

HEARINGS

BEFORE THE

SUBCOMMITTEE ON
CENSUS AND POPULATION

OF THE

COMMITTEE ON
POST OFFICE AND CIVIL SERVICE
HOUSE OF REPRESENTATIVES

NINETY-NINTH CONGRESS

SECOND SESSION

MAY 1 AND 15, 1986

Serial No. 99–48

Printed for the use of the Committee on Post Office and Civil Service

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

U.S. GOVERNMENT PRINTING OFFICE
61-902 O WASHINGTON : 1986

For sale by the Superintendent of Documents, Congressional Sales Office
U.S. Government Printing Office, Washington, DC 20402

TM 870 202

2

COMMITTEE ON POST OFFICE AND CIVIL SERVICE

WILLIAM D. FORD, Michigan, *Chairman*

WILLIAM (BILL) CLAY, Missouri
PATRICIA SCHROEDER, Colorado
STEPHEN J. SOLARZ, New York
ROBERT GARCIA, New York
MICKEY LELAND, Texas
GUS YATRON, Pennsylvania
MARY ROSE OAKAR, Ohio
GERRY SIKORSKI, Minnesota
FRANK McCLOSKEY, Indiana
GARY L. ACKERMAN, New York
MERVYN M. DYMALLY, California
RON DE LUGO, Virgin Islands
MORRIS K. UDALL, Arizona

GENE TAYLOR, Missouri
BENJAMIN A. GILMAN, New York
CHARLES PASHAYAN, JR., California
FRANK HORTON, New York
JOHN T. MYERS, Indiana
DON YOUNG, Alaska
JAMES V. HANSEN, Utah
DAN BURTON, Indiana

TOM DEYULIA, *Staff Director*
ROBERT E. LOCKHART, *General Counsel*
PATRICIA F. RISSLER, *Deputy Staff Director and Chief Clerk*
JOSEPH A. FISHER, *Minority Staff Director*

SUBCOMMITTEE ON CENSUS AND POPULATION

ROBERT GARCIA, New York, *Chairman*

MARY ROSE OAKAR, Ohio
GARY L. ACKERMAN, New York

JAMES V. HANSEN, Utah
JOHN T. MYERS, Indiana

LILLIAN FERNANDEZ, *Subcommittee Staff Director*

CONTENTS

May 1, 1986

PROCESSING PROCEDURES FOR THE 1990 DECENNIAL CENSUS

Statement of: Page

John G. Keane, Director, Bureau of the Census, accompanied by Peter A. Bounpane, Assistant Director for Demographic Censuses 2

Daniel G. Horvitz, executive vice president, Research Triangle Institute..... 36

William Eddy, department of statistics, Carnegie Mellon University 56

Judith S. Rowe, associate director for research services, Princeton University Computing Center..... 69

Benjamin F. King, director, survey methods, Educational Testing Service . 86

Stephen E. Fienberg, Maurice Faulk Professor of Statistics and Social Science, Carnegie Mellon University..... 98

May 15, 1986

CENSUS QUESTIONNAIRE AND AUTOMATION

Statement of:

Susan Miskura, Chief of Decennial Planning Division, U.S. Bureau of the Census..... 112

Gene Dodaro, Associate Director, General Government Division, U.S. General Accounting Office..... 150

Gail Franke, vice president, Federal Government marketing, National Computer Systems..... 183

4

PROCESSING PROCEDURES FOR 1990 DECENNIAL CENSUS

THURSDAY, MAY 1, 1986

House of Representatives,
Subcommittee on Census and Population,
Committee on Post Office and Civil Service,
Washington, DC.

The subcommittee met, pursuant to call, at 10:12 a.m., in room 311, Cannon House Office Building, Hon. Mary Rose Oakar presiding.

Ms. **Oakar**. Good morning and welcome to the Subcommittee on Census and Population hearing on the processing procedures for the 1990 decennial census. Unfortunately, Congressman Garcia, the chairman of this subcommittee, which I am proud to be a member of, must address some urgent matters in his district and regrets he is not able to be here. He does apologize, but I have offered to fill in, although no one really can, for Congressman Garcia and I will be chairing at least part of the hearing and hopefully Congressman Ackerman will take over. So we are delighted to be here.

I would like to submit Congressman Garcia's statement for the record.

[The statement of Robert Garcia follows:]

Opening Statement of Hon. Robert Garcia

Good morning and welcome to our hearing on processing the 1990 Decennial Census. We have decided to focus our hearing today on processing the census because it is critically important to the success of the decennial census.

The 1990 Census will be our nation's bicentennial census, and since 1790 the census has assisted our nation in attaining our democratic goals. The success of the census is central to the progress of our people and our economy. Because processing the census is linked to the success of the census, it is important to make sure that the Bureau utilizes the best automation technology available. In times of fiscal constraints, it is more than ever important to find ways to run our government programs more efficiently, including the decennial census.

In April of 1985 we held the first hearing during this Congress on the 1990 Decennial. At that hearing the General Accounting Office raised concerns that the Census Bureau is not making timely decisions regarding automating the census. Similar concerns were expressed at another hearing by the Inspector General of the Commerce Department. Now that it is a year later, I am interested in finding out the progress that has been made in the Census Bureau's decision making on processing the census.

Today, we have invited experts from the academic and private sectors to react to the Census Bureau's achievements and plans. These experts have been selected with assistance from the Committee on National Statistics of the National Academy of Sciences.

We will first hear from the Census Bureau on the status of their plans and activities in four major areas of processing: (1) address list compilation; (2) design of automated check-in procedure; (3) editing and processing the census forms; (4) and processing plans for data dissemination. Then we will hear evaluations of the plans from the experts in the four areas.

Ms. Oakar. The hearing was called to give the subcommittee an opportunity to learn the status of the Census Bureau's plans for processing the 1990 census. The Bureau provided the subcommittee with detailed reports on its plans, consisting largely of preliminary and internal documents, and these were referred to a panel of experts suggested by the National Academy of Sciences' Committee on National Statistics. The experts will also testify at the hearing.

The chairman believed that the subcommittee needed to be reassured that the Bureau formulated a clear plan of action for this vital part of the census. Critics are concerned that the Bureau is not considering a wide enough range of alternatives or settling upon a coherent management plan, and staff have been monitoring the process at the Census Bureau and keeping members appraised. The Bureau held planning meetings in October, but it is not clear how fully implemented those plans are. Due to the long lead times required to set up census procedures, the chairman was concerned that the Bureau not lose the momentum that it had. By holding the hearing it was hoped that the subcommittee could provide an occasion for the Bureau to explain how it will process the census by relying on outside, nonpartisan experts for the detailed reviews.

The subcommittee hopes to focus attention on the need for current action and away from partisan considerations. The cost of travel is not being reimbursed by the Government for the experts who have agreed to appear at their own expense, and we are very grateful for that. We are very delighted to have the Director of the Census Bureau here, Dr. John G. Keane. He will be accompanied by his Assistant Director, and we are very happy to have you, Doctor. Thank you very much for being here and, again, I know that my colleagues and Chairman Garcia welcome you and we look forward to your testimony.

STATEMENT OF JOHN G. KEANE, DIRECTOR, BUREAU OF THE CENSUS, ACCOMPANIED BY PETER A. BOUNPANE, ASSISTANT DIRECTOR FOR DEMOGRAPHIC CENSUSES

Mr. Keane. Good morning, Madam Chair, and thank you. We appreciate the opportunity to brief the subcommittee on our plans for the 1990 census. My full statement for the record has been submitted, so my oral comments will take the form of an overview.

As the subcommittee requested, I will discuss four crucial topics in planning the 1990 census. Those four are: one, concurrent processing; two, address list compilation; three, automated address control file and automated check-in; and, four, data products and their dissemination.

Going to one, concurrent processing, and by that just so we will all be clear on it, we mean questionnaire data conversion that occurs concurrently with questionnaire collection. We want to begin automated data processing of the 1990 census 5 to 7 months earlier than for the 1980 census. In 1980 the conversion of the data to machine readable form did not begin until all 409 district offices closed and shipped all questionnaires to our then three processing centers. So in 1980 it was a sequential process as opposed to a concurrent process which we are proposing and planning for 1990. The advantage of a concurrent process, some of the main ones are these: It will allow more time for review and correction of the data; it will enable the computer to assist in certain operations; and it will give us an earlier start that will help us meet our goal of disseminating data products in a more timely fashion.

The issues involved really boil down to two. Planning for the concurrent processing of the 1990 census has posed these two major questions: Where it would be done and how it would be done? The where issue involves the number of processing offices and the degree of centralization or decentralization. The how issue involves the technology we will use to convert questionnaire data into computer readable form.

In addition to these two major questions, we have to answer numerous related questions. As to the planning, we have been involved in extensive review of these issues over the last several months. For example, we held a major conference to discuss them in October 1985. That was referred to as the Decennial Census Decision Conference, and the subcommittee staff was represented at that conference.

Our discussions at this conference made clear what further information we would need to make a final decision; and since then the staff have prepared what we call action plans to analyze the benefits and the risks of various approaches identified. Senior staff completed the review of the first wave of these action plans earlier this week. Based on this review we have made some decisions on concurrent processing issues. At the conference in October we defined the basic structure of processing and collection offices. The action plans then analyzed this structure and found that several of its features were problematic. For example, the key entry workload would have required more offices, equipment, and staff than we believe we can manage. Developing two primary data conversion systems, that is, the key entry and FOSDIC, would be a strain on resources.

The required transfer of information between the processing and collection offices, especially for hard to enumerate areas, threatened delays in the start of the nonresponse followup. So we have reached decisions that will solve these problems. We have reduced workloads in certain key entry operations such as surname and full name keying. No data keying is planned. This will allow us to keep the number of processing offices down to a manageable number, down to about 12. We have agreed to use FOSDIC as the primary data conversion system for the 48 contiguous States. We have not yet reached decisions for Alaska, Hawaii, Puerto Rico, and the outlying areas.

Finally, we have agreed to configure the operations differently in areas that are hard to enumerate as opposed to the rest of the country. For the hard to enumerate areas, respondents will mail questionnaires back to the processing offices. There they will be automatically checked in, converted to computer readable form and computer editing. Clerks will prepare edits or route them for telephone followup. Questionnaires that need personal visit followup will be sent to the collection offices. The collection offices will conduct nonresponse and failed edit followup. For the rest of the country respondents will mail responses to the collection offices, and they will be automatically checked in. There will be a clerical telephone edit and personal visit followup as necessary.

As each questionnaire passes edit, the collection office will ship it on on a flow basis to the processing office for data conversion. The collection office will, of course, be responsible for nonresponse followup.

We believe these decisions represent the best balance of staffing, equipment, and workload considerations between the processing and collection offices. We avoid large staffing requirements in the processing offices by doing questionnaire check-in and edit for most of the country in the collection offices. For hard to enumerate areas, we will have the benefit of the automated edit. Also for hard to enumerate areas, we will not need large numbers of clerical workers. We can, therefore, concentrate on hiring followup enumerators. These decisions allow us to meet our goal of concurrent processing for the entire country.

Now that we have defined this basic office configuration, we will continue to work on the detailed plans for implementation over the next several months. We will make the final decisions by December of this year.

Two, address list compilation. We reported to the subcommittee on this topic at a March 13 hearing, so I will be very brief now. In 1984 we conducted an address list compilation test. The purpose of it was to test different ways to compile address lists. Based on the results of this test, we have determined our methods for compiling lists for both urban and rural areas. For urban areas we will use commercial vendor lists as the initial source. The test showed that the 1980 list could be a viable alternative or supplement to the vendor lists where the latter do not exist or where they are of suspect quality.

Currently we are developing criteria to evaluate, select, and augment the vendor list. For rural areas we will, again, have census enumerators create the initial list from scratch. We call this function free listing. For both urban and rural areas, we will conduct update operations to improve the list. The U.S. Postal Service will assist us in some of these updates.

In our test census we are refining our procedures. These efforts are described in some detail in my written testimony. Because of this extensive testing and fine tunning, we are confident that our mailing list for 1990 will be as good or better than those for 1980.

The third area is automated address control file and automated questionnaire check-in, and these are so interrelated we present them as one area. Now I shall discuss how we are going to use an automated address file to provide greater control over the census process and replace some large-scale clerical operations. In the 1980 census we did not have an automated address control file. Changes to the address registers resulting from district office operations were entered in pencil. Questionnaire check-in was done manually and involved many clerks and much time. Clerks also had to prepare separate address registers for nonresponse followup.

The development of an automated address control file will be one of the major automation advances since the 1980 census. We can key in changes to the lists and thus automatically update the file.

We can put bar code labels on the questionnaires and automatically check them in using laser sorters or wands. We can quickly identify addresses for which questionnaires have not been returned. The computer can print out lists of addresses for use by the follow-up enumerators. If we determine that it is cost-effective, we could send reminder notices to nonresponding addresses automatically, as we tested in Tampa. We are continuing to evaluate the cost-effectiveness of reminder notices.

Now to the final area, data products. The final topic I shall discuss is data products. It is fitting that this is our closing topic in a way. Making data available to policy makers and the public truly is the keystone of a successful census. We have been consulting with a broad array of data users in formulating our plans for the 1990 census data products: the National Conference of State Legislatures; special Census Bureau planning conferences; local public meetings, 65, at least one in a State; State data centers and State government agencies; regional meetings with American Indians; community meetings with Hispanic, black, Asian and Pacific Islander groups; other Federal agencies and, of course, this subcommittee staff.

During April and May of this year we are conducting a series of 10 product planning meetings around the country. In the fall of the year we will hold a conference to present our final report on the results of those meetings. At this time we are proposing that computer summary tapes, printed reports and microfiche will be the primary means for distributing 1990 census data, as they were for 1980.

We are also evaluating other dissemination media for making data available. These media include flexible diskettes for microcomputers, laser disks and on-line data systems. We are considering a number of changes in the format, the length and the sequencing of 1990 data products. These are intended to help us meet our goal of providing quality products with a minimum of delay.

We are discussing these proposals with data users at our product planning meetings. Our schedule calls for preparing the final program descriptions for both 100 percent and sample products by the fall of 1986. We plan to prepare the detailed specifications for most of these products during 1987.

In conclusion, each of the four issues I have discussed today raise many related complex issues, but we have made much progress in resolving outstanding issues and in making important changes since 1980. We are building a consensus plan, I might add a coherent plan in terms of our opening reference, that will lead to success in our collection procedures, in automated census processes and in the timely dissemination of useful data products.

Now I and my colleagues look forward to comments and suggestions from other witnesses assembled today.

[The statement of John G. Keane and his response to written questions follow:]

STATEMENT OF THE DIRECTOR OF THE BUREAU OF THE CENSUS

John G. Keane

Before the Subcommittee on Census and Population
Post Office and Civil Service Committee
U.S. House of Representatives
May 1, 1986

Introduction

Mr. Chairman, thank you for this opportunity to brief the Subcommittee on plans for the 1990 census. As you requested, I will discuss four topics that are crucial areas in planning the 1990 census. The topics are: (1) address list compilation, (2) automated address control file and automated check-in, (3) concurrent processing, and (4) data products and their dissemination.

We are at a very important point in planning the 1990 census. We must make a number of key decisions in the next few months that will largely determine how the 1990 census will be taken and processed. We welcome this opportunity to discuss our plans with you and look forward to the comments and advice of Subcommittee members and the expert witnesses who will be testifying today.

Address List Compilation

Mr. Chairman, the first topic I will talk about is address list compilation, which we discussed at the March 13 hearing before this Subcommittee.

Address lists that are as complete and accurate as possible are essential if we are going to conduct a good census using the mail-out/mail-back method. We use address lists to control the enumeration by mailing questionnaires to each housing unit on the lists and monitoring the mail returns to determine whether a questionnaire has been returned for a particular unit. Once a housing unit is included in our address lists, we stand an excellent chance of completing the enumeration of that unit and its inhabitants.

Address lists are not our only concern in taking a good census. We must also devise procedures to assure complete within-household coverage, to handle the enumeration of group quarters population, and to count those persons who have no usual living quarters. But, by compiling good address lists we go a long way toward having a successful census.

Since the mid-1960's, we have conducted several test censuses in which we examined address list compilation procedures, and we have compiled addresses for the last two decennial censuses. This experience has shown us that no single source of addresses is complete enough to meet our stringent requirements. Also, address lists become out-dated quickly. Therefore, we perform several updates to verify the accuracy and completeness of the census mailing lists.

In 1980, for urban mail-census areas, we first purchased address lists from commercial vendors. These lists were relatively inexpensive ($780,000 for 48.6 million addresses). They generally provided good coverage, but to improve them, we subjected the lists to four update operations before using them to mail questionnaires. The U.S. Postal Service (USPS) conducted three of these checks and our own enumerators conducted one.

We called the first USPS check the "advance post office check" because it was conducted about 10 months prior to Census Day (May-June 1979). The second check, conducted in early March 1980, was called the "casing" check, because mail carriers sorted census questionnaires into their cases but did not deliver them. We called the third USPS check the "time of delivery" check because it was conducted on March 28, 1980, when mail carriers delivered the census questionnaires. During each of these three checks, the USPS reported addresses missing from our lists to us, deleted addresses that were nonexistent or were businesses, and made corrections to addresses.

We called the update operation conducted by our own personnel "precanvass." In February 1980, our enumerators systematically canvassed assigned areas, updated the address lists, and verified the geographic locations of addresses.

For those mail-census areas where commercial lists were not available (generally, the more rural areas of the country), we hired enumerators to list and identify the geographic locations of addresses from scratch. We called this operation, which took place from March to October 1979, "prelist." We also subjected prelist addresses to the casing and time-of-delivery updates by the USPS. In addition, for selected areas, we added a canvassing operation after Census Day to identify any missed units.

As a result of all these overlapping checks, we believe the address lists were as complete as reasonably possible by the time we mailed the questionnaires. We continued to check on the completeness of our housing inventory in later census operations.

Given the importance of address lists in taking the census, it is appropriate that our first special test for the 1990 census was the 1984 Address List Compilation Test (ALCT). Even though we believe that the methods used to prepare the 1980 mailing list were successful, we wanted to test alternative address list sources to try to improve the accuracy and lower the costs of our methodology. A 1982 report by the General Accounting Office had suggested that we investigate the use of a mailing list developed by the USPS.

We designed the 1984 ALCT to evaluate several ways to develop address lists for both urban and rural mail areas. The overall evaluation compared the relative cost and the relative accuracy of various address list compilation sources as augmented by various update methods.

In the urban test sites (Bridgeport and Hartford, Connecticut) we compared three initial list sources: (1) the USPS, (2) a commercial vendor, and (3) the final list of addresses from the 1980 census. We updated each of these lists using our "precanvass" procedure and a USPS casing check.

Now, I will discuss the results of the ALCT for the urban areas. From the standpoint of coverage, the ALCT results do not rule out any of the list compilation techniques. However, the 1980 and USPS lists were more expensive than the vendor lists. Also, changing to a USPS-created list on a national scale would pose significant planning, control, and operational risks. Finally, the relative success of the 1980 list approach must be tempered by the fact that there had not been much change in the test areas since 1980. It is not reasonable to assume this approach would do as well in 1990 for high growth areas.

Based on these findings, we have decided to use vendor lists as the primary address list source in urban areas in 1990. The test also showed that the 1980 lists are an acceptable alternative or supplement to vendor lists in areas where no vendor has a list or the vendor list is of suspect quality. Thus, we will consider selective use of the 1980 lists. We are developing criteria for evaluating, selecting, and augmenting the vendor lists we receive in response to our request for proposals.

For the rural sites in the ALCT (Hardin County, Texas and Gordon and Murray Counties, Georgia) we tested two initial list sources: (1) the USPS, and (2) a Census Bureau prelist operation. We also used the same two update techniques as in the urban areas: a USPS casing check and a precanvass by census enumerators.

Again, from the standpoint of coverage, the ALCT results do not rule out the USPS list. There also do not appear to be any significant differences in cost between the two methodologies. However, the USPS had problems assigning correct geographic codes to addresses and marking housing unit locations on census maps (which is essential for field followup purposes). Furthermore, there are the same risks mentioned above for urban areas if we were to convert to the use of USPS-developed lists in rural areas on a national scale.

Based on these factors, we have decided to use the prelist methodology to create the initial address list for rural areas in 1990.

Despite the decision not to use the USPS as a source of our initial address lists for the 1990 census, the USPS still will play a crucial role in the 1990 census. The USPS will conduct the three coverage checks of our address lists, deliver and collect the questionnaires, and, as I will discuss below, sort the returned questionnaires for us.

As a result of the 1984 ALCT, we have determined our approaches for acquiring initial lists for the 1990 census. I will close this discussion of address list compilation by reviewing briefly some of the additional tests we are conducting that relate to address list methodology.

In the 1986 test census in East Central Mississippi, we are refining our prelist procedures. We have incorporated an additional USPS check of the prelist addresses--an advance post office check. In 1980, this postal check was only done for the vendor lists in urban areas. We also are looking at rural delivery methodology. Following the prelist and advance post office check, the Mississippi test site was split into two panels. In one panel, census enumerators conducted a precanvass operation followed by USPS delivery of questionnaires. In the other panel, census enumerators delivered questionnaires and updated the address lists at the same time. This new update/leave operation is being tested because in the past there have been difficulties with our ability to obtain adequate mailing addresses for certain rural areas of the country. We will compare the coverage and cost between these two methods, the operational feasibility of each, and problems in coordinating two types of delivery.

In our 1985 test censuses in Jersey City, New Jersey and Tampa, Florida, and in our 1986 test census in Central Los Angeles County we tested and evaluated refinements to the precanvass procedures. First, we scheduled precanvass about 2 months earlier than in 1980. Scheduling precanvass earlier enables us to incorporate changes so as to have a more accurate and complete housing unit inventory for review during the USPS casing and time-of-delivery checks. Second, we had the enumerators do a unit-by-unit canvass in each multiunit building. In 1980, the enumerator only verified the number of units in multiunit structures from the manager and did not actually canvass within the building unless the reported number exceeded our counts. Preliminary results from the 1985 test in Jersey City show that 9 percent of the apartment designations were updated and improved as a result of this new procedure. This should help to improve the accuracy of delivery and followup operations, particularly in those hard-to-enumerate areas where there are many multiunit structures.

This is just a sampling of the 1985 and 1986 tests related to our mailing lists. We will continue to refine our procedures in the 1987 test census. To sum up: At this point, we have determined the sources for initial lists (one of our major milestones in 1990 census planning) and are determining the most cost-effective set of update operations to those lists. We expect our mailing lists for 1990 to be as good, or better, than those for 1980.

Automated Address Control File
And
Automated Questionnaire Check-In

I will now discuss how we are going to use an automated address file to provide greater control over the census process and to replace some large-scale clerical operations (such as questionnaire check-in) with automated processes.

The development of an automated address control file is one of the major automation advances since the 1980 census. A description of the labor-intensive 1980 census operations will be helpful in understanding how an automated address control file will improve census procedures in 1990.

In 1980, although the initial file of addresses was computerized, the district offices received only paper lists of the addresses--one or more address "registers" containing the address for each housing unit in the enumeration district (ED). Clerks made manual changes to the address registers as a result of operations such as precanvass and the USPS casing and time-of-delivery checks. The changes included adding, deleting, and correcting addresses, as well as moving them from one ED or census block to another. These update operations were labor-intensive and, as with any large-scale clerical operation, subject to error.

We sorted and checked-in returned questionnaires in a manual fashion. When householders returned their questionnaires by mail in 1980 (about 70 million did so), our clerks in the district offices first sorted the forms by hand to the ED level and then placed them in serial number order within ED. Then the clerks matched the questionnaires, one at a time, to the address register for the appropriate ED. When they found the serial number and corresponding address in the book, they recorded the check-in date and other pertinent data. This operation required dozens of clerks in each of the 409 district offices. Many district offices took longer than the 2 weeks allotted, delaying the start of followup operations.

For followup of nonrespondents in 1980, we gave the address registers to the enumerators and kept a copy in the district office. (The addresses were printed on two-part paper and two books were created by separating the pages and reassembling them.) The address books contained all the original computer-printed addresses and all hand-entered changes, and contained every address--both for those households that responded to the census and those for which we received no response.

With an automated address file, we can key in changes and, thus, automatically update the file. We can use bar-code technology for computer check-in of the questionnaires. As a result, it will be easier to identify quickly the addresses for which questionnaires have not been returned. If we determine it is cost-effective, we could send reminder notices to those addresses automatically, possibly reducing further the number of nonresponding housing units to which we need to send enumerators. It also should be noted that concurrent processing of census data, which I will discuss next, would not be feasible without an automated address control file and automated questionnaire check-in.

On the other hand, there are risks involved in automating the address control file. Most importantly, having an automated address control file to support field operations will increase our requirements for computer hardware and skilled personnel to operate that hardware in the processing centers at the peak of operations.

In our 1985 test censuses in Jersey City, New Jersey, and Tampa, Florida, we successfully implemented an automated address control file, automated check-in, and the use of reminder cards.

We are building on our 1985 experiences in the test censuses this year. Census Day was March 16 for our 1986 test censuses in Central Los Angeles County, California, and East Central Mississippi. In Los Angeles County, we have a collection office in Bell but we are processing the census data at a separate site in Laguna Niguel. Householders mailed their questionnaires directly to the processing office in Laguna Niguel. The USPS sorted the questionnaires for us by form type (short or long). An additional sort was performed in the processing center for single- and multiunit addresses. (The questionnaires for multiunit addresses went to a keying operation where the householders' surnames were keyed in. This is necessary for nonresponse followup.)

We had imprinted each questionnaire with a bar code that contains a unique identification number. We ran the returned questionnaires through a multiple-pocket laser sorter that performed the necessary sorts not already done by the USPS and simultaneously read each questionnaire's bar code, recording the identification number onto a computer tape attached to the sorting equipment. We used this tape to update the address control file, and the addresses for which questionnaires were returned were noted on the address control file along with the date of check-in. For various reasons, the laser sorter could not read all bar codes. In these cases, clerks used hand-held wand readers to record the data, and if the wands also were unable to pick up the codes, we relied on keying as a back up.

Using this automated system, we could determine from the address control file whether a questionnaire had been checked in for a particular address. We generated lists of nonresponse addresses and mailed reminder cards to them. On April 3, 18 days after Census Day, we generated the lists of nonresponse addresses to be followed up by enumerators. The followup lists contained only addresses for which questionnaires were not received. For multiunit addresses, all addresses were listed (along with the names of the householders and other information for responding units) so the enumerators could resolve possible apartment mixups. Units for which questionnaires had been returned are not being visited, except when necessary to clear up apartment mixups. Apartment mixups occur in buildings without clear unit designations or when the mail carrier gives the questionnaire designated for Apt. 1 to Apt. 2, etc. Finding ways to solve apartment mixups is an important part of our plans to improve the census in hard-to-enumerate areas.

In Mississippi, where we combined the collection and processing offices, the principle of automated check-in was much the same as in the urban test site (i.e., use of bar-code technology), but there were some differences. The USPS did not perform any questionnaire sorts for us, nor were laser sorters used. All sorts were done manually by processing office staff. We used pen-shaped wands connected to micro computers to check-in the questionnaires. Clerks moved a wand over the bar codes to read the identification numbers. They keyed in the numbers if they could not be read by wand.

One of the major decisions we must make with regard to check-in is which technology to use in which type of processing office. This issue is intimately tied to the larger issue of office configuration (number and type of processing offices). It is estimated that one laser sorter can process about 11 times as many questionnaires as one wand station (10,000 per hour vs. 900) but the laser sorter could cost about 50 times more ($250,000 vs. $5,000). The laser sorters would require more maintenance per machine and more skilled personnel. On the other hand, using wands more extensively means more wand stations and production clerks would be required. We also must consider potential use of the equipment after the 1990 census. (Micro computers used in wanding would have more use after 1990 than laser sorters.)

Another issue we have examined is the extent of USPS involvement in questionnaire sorting. We already have reached an agreement with the USPS under which they will sort questionnaires for us by district office, form type, and address type (single- or multiunit).

Final decisions on the issues related to the automated address control file and automated check-in are scheduled to be made by September 1986. In the meantime, working groups are preparing analyses that examine the various options, and we must evaluate our experiences in the test censuses in Los Angeles County and Mississippi, where questionnaire check-in recently occurred.

Concurrent Processing

Now, I will turn to our third topic, concurrent processing. The essence of concurrent processing is that we want to begin the conversion of questionnaire data to machine-readable form concurrently with the questionnaire collection operation.

In 1980, while we conducted some basic questionnaire processing in our temporary district offices, the conversion of data to machine-readable form did not begin until after the district offices completed all enumeration, edits, and followups and shipped all questionnaires to one of three automated processing centers. This was a sequential process. An earlier start in 1990 (5-7 months ahead of the 1980 schedule) will allow more time for review and correction and will enable the computer to assist in certain census operations. It also will contribute to the early identification of enumeration problems. Also, by converting questionnaire data to machine-readable form sooner, we can minimize the potential for losing data when original questionnaires are accidentally damaged or destroyed. Finally, and perhaps most importantly, it will help us meet our goal of disseminating data products in a more timely fashion.

One of the operations which could be done by computer is editing of the questionnaires. In 1980, certain basic edits for completeness and coverage were done clerically in the field offices; then, later, the computer completed other edits (such as for consistency of the data). We have been examining whether we can automate the edits done clerically in 1980. Computer edits would be more accurate than clerical edits. Staffing might be reduced somewhat, although we would still need clerks to control and review the questionnaires that fail edit to see if they can be clerically "repaired." We also need clerks to make telephone calls or personal visits to those respondents whose questionnaires cannot be repaired. While automated editing has some important advantages it has drawbacks as well. Perhaps the most critical are the large number of review and telephone clerks that would be needed in the processing centers, and the need to convert the data a second time to computer-readable form and assure that it replaces the first data. We are weighing these pros and cons to see if we want to do automated editing as part of concurrent processing.

Planning for concurrent processing in the 1990 census has posed two major questions: Where and how would it be done? The "where" issue involves the number of processing offices and the degree of centralization or decentralization. In 1980, when we processed the census questionnaires sequentially, we had three processing centers. With concurrent processing having so few centers probably would not be feasible, primarily because of the need to move materials quickly between processing and collection offices. Greater centralization of processing activities also places greater staffing burdens on us, i.e., the need to hire more employees in one employment area. We must weigh these concerns against problems related to decentralization, such as the need for more hardware to service a greater number of locations and the difficulties of controlling and supporting many processing offices.

The "how" issue involves the technology we will use to convert questionnaire data into a computer-readable format. In the 1980 census, we used the FOSDIC technology to do this. FOSDIC stands for Film Optical Sensing Device for Input to Computer. The complete data-conversion system consists of high-speed cameras that film the questionnaires, film developers to process the raw film into rolls of microfilm, and the FOSDIC machines that read the data from microfilm onto computer tape. We call the system FACT, which stands for FOSDIC and Automated Camera Technology. This system worked very well in the 1980 census, and also was used in the 1960 and 1970 censuses. We are currently evaluating the use of the fact system for the 1990 census, as well as, considering an alternate primary data conversion technology--data keying. Even if keying is not used as a primary data conversion technology, we will use it for entering some of the handwritten data on the questionnaires into computer-readable form.

In addition to the two major questions of where and how to do concurrent processing, we have to answer numerous other related questions. We need to make our major decisions on processing methodology by September

1986 and begin procuring required equipment by early 1987 (some procurements will begin sooner).

As a start to answering all of these questions we conducted the Decennial Census Decision Conference in October 1985.

At this conference, we decided that we should implement automated processing earlier for the 1990 census than for 1980--so that it occurs concurrently with field operations. We also agreed on a general office configuration--not as a final decision, but as a framework to help us focus our future planning efforts. Under the proposed configuration, we would locate large "host" processing offices in metropolitan areas to serve several district offices. In more rural areas of the country, we would combine the district and processing offices. We would conduct both automated processing activities and field follow-up from the same office.

We also discussed extensively the issue of data conversion technology since it is intertwined with the issue of office configuration, but a decision on this issue was not a goal of the conference.

As we discussed these issues at the conference, we found that we did not have all the information we needed to reach final decisions on the issues related to concurrent processing. After the Decennial Census Decision Conference, we established working groups to analyze these issues and to prepare "action plans". Each action plan is designed to examine and answer specific questions about the data processing systems necessary to support the 1990 census and to analyze the benefits and risks of various approaches in terms of office structure, timing, staffing, costs, management and coordination, quality, technical support, etc.

The action plans are being completed in three waves. We have provided the Subcommittee with a complete list of action plans and the three key plans in the first wave. I want to emphasize that these reports represent internal Census Bureau staff analysis and recommendations: They were not thoroughly reviewed by senior Census Bureau management nor the Department of Commerce before release to the Subcommittee and do not represent official Census Bureau positions.

The set of issues we are addressing to make concurrent processing a reality are very complex. All of the action plans, including the other first-wave plans and later plans, are interrelated such that a change in one plan can dramatically affect other plans. (Third wave plans are perhaps less interrelated.)

We are still in the planning process on the issue of concurrent processing, but we have made progress in narrowing the options during the last 6 months. We have set up a process to analyze thoroughly the many complex and interrelated issues. I am confident this process will lead us to a decision that represents the best approach for the 1990 census.

We will evaluate the second and third wave action plans in September 1986. At that time, we will make final, detailed decisions on the office configuration, the distribution of equipment, and the other issues related to concurrent processing.

Data Products

The fourth and final topic I will discuss today is data products. It is fitting that this is our closing topic today, because producing and disseminating data products is the keystone to a successful census. All that comes before is geared toward producing accurate data in a timely manner. Each of the previous topics I have discussed today relate to these two goals: accuracy and timeliness.

I will begin by describing briefly our process for planning the 1990 data products, which has been underway for some time. In 1982, we began an internal appraisal of the 1980 census data dissemination program. In 1982-1983, the National Conference of State Legislatures (NCSL) surveyed state legislative officials regarding the 1980 census P. L. 94-171 (redistricting) computer tapes, microfiche, and paper printouts. In October 1983, the Census Bureau held a national conference of state officials and minority group members to discuss the P. L. 94-171 program and the NCSL survey results.

In April 1984, we held a National Geographic Areas Conference to examine our approach to the definition, delineation, and recognition of geographic areas in the 1980 census and how these activities should change for the 1990 census. In October and November of 1984 we held three Regional Geographic Areas Conferences to share the results of the National Conference with a wider audience and to gather additional information on these fundamental units for data tabulation.

In the period 1984-1985, we held 65 Local Public Meetings (at least one in each state); these meetings produced suggestions for planning 1990 data dissemination media. Also in this period, the State Data Centers, state government agencies, and regional and local planning organizations provided comments on their use of all 1980 data products and their recommendations for 1990. We held regional meetings with American Indians and Alaska Natives to obtain their suggestions and held community meetings with Hispanic, Black, and Asian and Pacific Islander groups to collect similar recommendations. We have also discussed our plans with professional organizations and census advisory committees. In the near future, we will be consulting with other Federal agencies to learn their needs for data products.

During April and May of this year we are conducting a series of product planning meetings around the country. In the fall of 1986 we will hold a conference to present our final report on the product meetings. Based on the thousands of recommendations we received in earlier meetings, we have developed proposals for 100-percent reports (those based on data we receive from all respondents), user tapes, and microfiche. We are discussing these proposals at the product planning meetings. Many of these proposals also relate to the sample products (those based on data we receive from a sample of respondents) and we will also be discussing these at the meetings. We will discuss sample products more fully at the fall meeting.

At this time, we propose that computer summary tapes, printed reports, and microfiche will be the primary means for distributing 1990 census data, as they were for 1980. Printed reports are an essential medium of data dissemination. While access to computers continues to grow, the Census Bureau is committed to making data available to all segments of the population. The printed reports, available through libraries, make this possible.

Microfiche is a compact and relatively inexpensive way of making a large amount of data available in "eye-readable" form. Thus, we now plan to make selected summary tape files and block statistics available on microfiche. In 1970 and earlier censuses, block statistics were issued in printed reports. In 1980, however, the number of blocks covered in the series increased to 2.5 million and printed reports would have been too costly to produce and too cumbersome to store. So we issued the statistics on microfiche. For 1990, we may have as many as 12 to 15 million blocks. Printed reports for these would be even more costly and cumbersome than in 1980, so we will again plan to use microfiche. We are looking at ways to make the microfiche easier to use. One option is to use larger type, but the trade-off is that more fiche would be required to hold the data. The number of fiche also will increase due to the greater number of blocks.

We are also evaluating other dissemination media for making data available. These media include flexible diskettes for microcomputers, laser disks, and on-line data systems.

Since 1984, the Census Bureau has sold selected data products on flexible diskettes. Given the large quantities of data on the 1980 census summary tape files and the limited capacity of diskettes, we have not considered preparing diskettes containing such data. For example, the 1980 census Summary Tape File 3 data for Montana are contained on one reel of magnetic computer tape but would require over 100 diskettes. We do plan to explore the possibility of downloading small subsets of summary tape files onto diskettes using test data tapes from the 1988 dress rehearsal census. If users react positively to such diskettes, we may offer similar products containing 1990 census data. In addition, we are considering the feasibility of producing smaller sets of data on diskettes on request.

Laser disks, similar to the compact disks being sold for home audio systems, offer considerable potential for data dissemination. Although small in size, they provide tremendous storage capacity. One laser disk will hold the contents of about four high-density computer tapes. We are currently evaluating laser disks to see if they are a viable option for our 1987 agriculture and economic censuses and the 1990 decennial census.

We initiated an on-line database service in 1984. Called CENDATA, it is available to the public through private sector database firms. At present, CENDATA is relatively small, containing mostly current summary data from ongoing surveys and product information. For 1990, we expect to expand the capabilities of CENDATA to provide graphics, an interactive mode for inquiry and order handling, and substantial amounts of additional timely summary data. Even with expansion, however, only a small part of 1990 census data would be available through this system.

Now, having described our thoughts on dissemination media, I will discuss a few of the other product issues we are addressing in our product planning meetings.

We are proposing a number of changes in the format and sequencing of 1990 data products to help us meet our goal of providing quality products with a minimum of delay. For example, we are proposing that the earliest products from the census exclude those statistical areas that will be defined on the basis of 1990 census results. These nongovernmental entities are metropolitan statistical areas and urbanized areas. These areas will appear in later products for 1990.

For the same reasons, we are also proposing that we limit the amount of historic data in the initial release and concentrate most historical data in a special series to be issued later.

Some users have expressed a preference for more reports that combine population and housing data. We propose this for all the 100-percent data products. Because products based on sample data are larger than the 100-percent products, it may not be possible to follow the combined approach.

Many 1980 census printed reports were much larger than their 1970 counterparts, at least partially because of the decision to publish much more data for race and Spanish origin. We published about as much detail for the major race groups (that is, White; Black; American Indian, Eskimo, and Aleut; and Asian and Pacific Islander) and for the Spanish-origin population as we did for the total population. In planning meetings we have held so far, such as the Local Public Meetings, some data users have told us that this level of detail is not necessary in the printed reports. In our current product planning meetings, we are asking data users whether the amount of data needed by race and Spanish origin could be reduced in the 1990 reports. Regardless of the amount of detail shown in the printed reports, we still plan to tabulate as much data for race and Spanish-origin groups as in 1980, and these data would be available in the summary tapes.

Finally, I will mention that we are considering eliminating both the computer summary files and the printed reports that show the most detailed cross-tabulations of population and housing data. Again, we are considering this based on user recommendations. These data would be available as special tabulations on a reimbursable basis. We also plan to develop public-use microdata samples that would allow users to do their own detailed cross-tabulations.

This is just a sampling of the many issues we must address in planning our 1990 census data products and we look forward to working with the Subcommittee further on these and other issues. Our tentative schedule calls for preparing the final program descriptions for both 100-percent and sample products by late 1986 and the detailed specifications for 100-percent and sample products during 1987.

Conclusion

Mr. Chairman, this concludes my testimony. In each of the four areas I have discussed today, we have faced or are still grappling with many complex issues. But we have made much progress in resolving outstanding issues and in making important changes since 1980. We are building a census plan that we believe will lead to success in our collection procedures, in automating census processes, and in disseminating useful data products in a timely manner. Now, I look forward to hearing the comments and suggestions of the other witnesses assembled here today.

Responses to Questions from
Subcommittee on Census and Population
to
Dr. John G. Keane
Director
Bureau of the Census
on
Processing Plans for the 1990 Decennial
May 1, 1986

QUESTION 1. In your testimony, you indicated that the Census Bureau was considering cutting back on the tabulations of the 1990 Census. In particular you said that you would produce fewer tabulations for Blacks and Hispanics. You know that I have long believed that the census is a very important source of information about the progress of the people and especially those people who need the most help. Following are a few questions about your decision:

> Exactly what tabulations are you thinking of taking out of the publications?
>
> How do you expect to serve the needs of the people who need those tabulations?

ANSWER: We have not made any decision on possible reductions but are exploring various approaches based on comments received at the local public meetings held in 1984 and 1985. Some data users at these meetings suggested that we reduce the amount of subjects or the number of smaller geographical areas shown for Spanish origin and race groups.

Since one of our primary concerns is to meet the major data needs for Spanish origin and race, we are seeking additional advice on this matter. We are raising this issue at a series of ten meetings on 1990 census data products being held in selected cities across the country. We plan to consult with the 1990 Advisory Committees on the Hispanic, Black, Asian and Pacific Islander, and American Indian and Alaska Native populations, and with ethnic leaders and community-based organizations. We also will keep this Subcommittee informed of developments on this issue and welcome your comments.

If printed data on race and Spanish origin were reduced, the data would still be available from the summary tape files. We expect that the amount of 1990 race and Spanish origin data available from the tapes should be about the same as in 1980. Also, to make data more accessible to all data users, we would consider expanding the amount of 1990 data on race and Spanish origin available on microfiche and other media.

As was done with the 1980 census data, the State Data Centers and their affiliates, upon request, would compile 1990 data on Spanish origin and race from the summary tape files and make the data available to the data users at a nominal cost.

QUESTION 2. When the National Academy of Sciences' study of plans for the 1990 Census was issued, our understanding was then that you would continue to support the work of this panel. Will the panel continue its work? Will you ask the NAS to prepare another followup report? Will you then provide full funding for this study?

We regard the NAS work as a very important nonpartisan source of information about the census. Has the Bureau agreed to continue funding it?

ANSWER: The National Academy of Sciences has made significant contributions toward the Census Bureau's establishment of research and operational priorities for the 1990 Decennial Census. These contributions are of particular importance to programs such as decennial census undercount and adjustment research.

We are contracting with the Academy to convene a special panel meeting in the late summer of 1986 to advise on the development of adjustment-related programs. The Academy will present a report of observations and recommendations to us in December 1986. Recommendations will help us develop the decennial census adjustment standards.

To ensure continued participation by the Academy, we are negotiating for a two-year extension of the existing contract. During the second year, this contract extension will provide for on-site consultation with Bureau of the Census research principals to ensure an opportunity for the exchange of information on a schedule compatible with our commitment to critical milestone dates.

QUESTION 3. I understand from your testimony that the Bureau has decided to use vendor lists as the main source for addresses to mail out the census forms. However, you will also be using the 1980 census lists in areas where the vendor lists are not considered to be very complete. Could you clarify this point? How will you select the areas of the country where you will use vendor lists and those where you will use 1980 census lists?

ANSWER: Vendor files will be evaluated to determine which are best for specific portions of the urban United States. As part of the evaluation, we will identify those urban areas where clear-cut deficiencies exist in the vendor files. For these areas we will consider the use of the 1980 census address list as a useful substitute or supplement to the vendor files. We are currently developing guidelines and criteria for evaluating, selecting, and augmenting purchased vendor files. We will provide the guidelines and criteria to the Subcommittee staff when they have been completed.

QUESTION 4. Many of your decisions relating to the processing of the census are based on the pretests. We have heard reports that the pretest in Los Angeles had to be curtailed due to poor response from the public. Could you please tell us what happened? How has it affected the test and what it will mean for the census?

ANSWER: We planned to conduct full-scale censuses in two offices in Central Los Angeles County--the North office serving 9 communities and the South office serving 12 communities. We also planned, as a contingency, to curtail some follow-up activities if mail-response rates were below expectations.

Because of the extremely low mail-response rate in Los Angeles, we determined that we could not complete the test procedures throughout the planned test area with available resources. We decided to reallocate available resources from the South office to complete the test in the North office. We decided to continue the test in the North office as opposed to the South office because the North office contained more test objectives and evaluations. Also, the North office is smaller and had a somewhat higher mail-response rate.

It is difficult to conduct a successful promotional campaign in a test census environment without the massive national publicity and visibility normally available during the decennial census. This problem is particularly severe when the test area is only a portion of the market served by the metropolitan mass media. For this reason we projected a mail response rate (45-50 percent) for the Los Angeles test sites much lower than it would be for that area in a national census.

The initial mail-response rate in Los Angeles was even lower than we expected. At the time we had to make our decision (March 27), the rate for the South office was 24 percent and for the North office, 31 percent. By April 3, when we printed the nonresponse follow-up listings the rates were 30 percent and 38 percent, respectively.

To help us evaluate the reasons for this low response rate, we surveyed people in both the North and South areas to determine their exposure and reaction to our promotion and community outreach activities. We have efforts underway to develop an effective national promotional campaign in 1990, and we will study what occurred in Los Angeles to suggest ways to improve these efforts.

Before our 1986 test census in Los Angeles, we had defined a number of objectives specifying what we planned to learn in conducting the test. Some of the new or modified procedures we were testing applied to both the North and South offices; a few were concentrated in the North office by design. We believe we will still be able to learn virtually all we had hoped to.

For those objectives that apply to operations that occur before Census Day, we will have data from both the North and South offices. For those that apply after Census Day we will still have enough good information to give us confidence in the conclusions we draw from the 1986 census of Los Angeles County.

QUESTION 5a. Among the documents you sent us to read there were a number that set out some of the issues that are related to processing the census. We greatly appreciate your sharing these with us even though they were preliminary. But they do raise a number of questions. How will the decisions you are making regarding the processing affect the kind and timing of the evaluation of the census coverage? When will we learn how complete the census was?

ANSWER: Planning for prompt coverage evaluation has been an integral part of overall census planning for 1990. This is most evident in our early commitment to pursue the dual strategy of (1) conducting the most complete census possible and (2) measuring and being prepared to adjust for coverage error on a timely basis. The requirements of the evaluation program have been carefully considered throughout the planning process, and all decisions, including those for processing, are thoroughly reviewed for their effects on the evaluation program.

We believe that several components of our processing plans for 1990 will significantly improve the accuracy and speed of our coverage evaluation program compared to 1980. The three most important components affecting coverage evaluation are: (1) automating the address control file, (2) converting questionnaire data to computer-readable forms earlier (concurrently with data collection), and (3) keying the names of persons in census blocks in or adjacent to areas where we will conduct the post-enumeration survey. The major effect of these automated processes is to allow us to begin earlier than in 1980 to match names from the post-enumeration survey to the census and to automate most of this matching (instead of doing it clerically as in 1980).

As we proceed to implement our recent processing decisions, we will determine the exact timing for estimating coverage. We believe it is essential that the results of the coverage evaluation program be as timely as possible.

QUESTION 5b. Have you decided to enter all of the names of the people into your computer? I understand that this has never been done before.

ANSWER: No. We will enter surnames of householders in multiunit structures for which a questionnaire was returned by mail. This automated operation replaces a clerical operation in the 1980 census, when clerks wrote surnames into the address registers. The surnames are needed by follow-up enumerators who may run into apartment "mixups" in multiunit structures. For example, the enumerator might visit Apt. 3G and find that Mr. Doe says he has returned a questionnaire. The enumerator can look at his listing and find that Mr. Doe returned a questionnaire but it was for Apt. 3K.

We will also enter full names for the relatively small sample of persons who live in the areas that will be covered by our coverage evaluation survey (the post-enumeration survey) or areas adjacent to those covered by the coverage evaluation survey. Names are needed to automate the matching of persons from the survey to the census. Clerically matching names was one of the major problems in the 1980 Post-Enumeraton Program.

Before the census, when we compile the address lists for the more rural areas, we will enter surnames for those addresses that are not street name and number (for example, rural route delivery addresses). We did this in 1980 also. For 1990, we will also have the ability to enter updates when the names of householders change. Names are an essential part of the mailing addresses in rural areas.

QUESTION 5c. How will you protect the privacy of the people whose names are in the computer?

ANSWER: We have established as one of our six major goals for the 1990 census the maintenance of the confidentiality of individual census information. This is especially a concern as we move to increased use of automation in the census. We will not introduce automated systems unless we can maintain the security of the data.

As in the past, we will make all employees aware of the Title 13, United States Code, confidentiality requirements imposed on all Census Bureau personnel. We will control the number of people with authorized access to confidential information and implement safeguards to prevent unauthorized access. Although we will have addresses and a small sample of names on automated files, we will store the names and addresses separately from the respondent-supplied questionnaire answers. And we will program our computerized systems to monitor against unusual or unauthorized searches of data files.

QUESTION 5d. Are you going to use the names to match the census returns with other government files? Why are you doing this? What will be its benefits and drawbacks?

ANSWER: We do not intend to match our census files to other government records to create a master data file of information about individuals. During the census-taking operations, however, as a method of identifying persons who were potentially missed, we are developing a program that will allow us to compare other files to the census. We will enumerate those persons that we verify were actually missed in the census. This program will allow us to use these other sources to improve the count of the population but only as protected under our guarantees of confidentiality.

QUESTION 6a. In one of the documents the Bureau sent us there was a discussion of how you might adjust the census returns to make up for an undercount. Now, I realize that this was just a staff paper but it does raise a number of questions. The paper presents a timetable for collecting information about the quality of the census coverage and reporting it to you. If I understand it correctly, it would require you to complete the fieldwork of the census in July. The fieldwork of the census in 1980 was not finished until well into the fall. Do you think you can make this time table?

ANSWER: Our goal will be to complete field work in as many offices as possible in July. The increased use of automation and improvements in census procedures and in hiring, retaining, and managing enumerators may give us a better chance than in 1980 of closing offices this early. As field offices complete the enumeration, data collection for the coverage measurement survey will begin. We will be able to start this data collection on a flow basis to allow for some office closings later than July and still complete all processing on a timely basis.

QUESTION 6b. The paper calls for keying lots of names of people. Do you think you will be able to do this?

ANSWER: Yes. We have decided only to key surnames for householders in multiunit structures who return their questionnaires by mail, a small sample of full names, and updates to names of householders in rural-route delivery areas. The full name keying will be done only for those blocks in or adjacent to areas where we are conducting the coverage evaluation survey. We anticipate no problems in performing this smaller keying workload in time to begin the coverage evaluation survey matching.

QUESTION 6c. The paper says, "Under the strategy adopted by the Bureau, the director is to review the results of both the census and the coverage measurement survey, compare them with the established standards and announce a decision on whether the census will be adjusted." (page 130)

In other words, the paper says that the Census Bureau director will decide whether to adjust the census after he or she knows what the effect of adjustment will be on the apportionment of the House of Representatives.

Is this the Bureau's current policy?

ANSWER: We will make our decision on whether to adjust the 1990 census figures after we know the results of our coverage measurement program and the quality of both the enumeration and the evaluation. Our major criterion will be whether we can improve the counts by adjusting them. While we may know the effect of adjustment on apportionment at that time, that will not be considered in making our decision. The Director will make a decision based on established standards that will be printed in the Federal Register before the 1990 census.

QUESTION 6d. How will you decide whether to adjust the census?

ANSWER: We will decide whether to adjust based on the results of our coverage measurement program and the results of the census. We will be looking at whether adjustment will improve the counts. We will review the results of both the census and the coverage measurement survey and compare them with established standards. The key to this strategy is reaching consensus on the standards before the census is taken. We will publish proposed standards in the Federal Register well in advance of the census.

QUESTION 6e. Will you be up here soon asking us to change section 195 of Title 13 to give you the authority to use sample data for the apportionment of the Congress?

ANSWER: Following the 1980 census, the courts reviewed whether Section 195 prohibits the use of any data based on a sample. In three cases, City of Philadelphia v. Klutznick, Carey v. Klutznick, and Young v. Klutznick, the district courts interpreted Section 195 to mean that sampling is prohibited only if it is the sole means of determining the population for apportionment. Based on these decisions we do not see any legal barrier to supplementing the count resulting from a full-scale enumeration by techniques that use sampling. If there is any ambiguity concerning that interpretation, the law may need to be clarified before any action.

QUESTION 6f. Now, we don't know who the Director of the Census will be in 1990. That largely depends on the outcome of the 1988 election. Assuming that someone else was the director and that he or she was appointed by a president of a different party from your own, would you trust him or her to make the decision about adjusting the counts after it was known what the effect on apportionment would be?

ANSWER: The plan for determining whether or not to adjust has been designed so as not to be dependent only upon the judgment of one individual. By developing criteria that are agreed upon in advance, we are removing the need to trust the judgment of one specific person and/or any concern about whether the effect upon apportionment is known to that individual. We believe that the best plan for a decision must be based upon definite knowledge about the results of our coverage evaluation program and the quality of both the census and the evaluation.

QUESTION 6g. Why wouldn't it be better to decide whether to adjust before the census is taken and avoid the difficulty of having to make a controversial decision under a great deal of pressure? Wouldn't this allow for a fuller debate on the topic and more careful planning of how it will or will not be done?

ANSWER: We will decide whether to adjust based on the results of our coverage measurement program and the results of the census. Thus, we cannot decide until after the census. Also, because there are statistical problems associated with the program to develop measures of the undercount, we must know how well it works, and at what geographic levels we might be able to perform an adjustment, before we can make a responsible judgment about whether or not adjustment will improve the census counts. There is ample time, however, for a full discussion and careful planning of the standards that will be used to make that judgment.

QUESTION 7. Regarding the costs of the processing that you are proposing for the census--will the equipment that you are purchasing be cost effective? How much money will the government save from the processing alternatives that you are considering?

ANSWER: Yes, we believe our equipment decisions will be cost-effective. We will only invest in automation that reduces costs or significantly improves the census. While we cannot know at this time whether a specific automation decision will save money, we believe our decisions will lead to greater efficiency and to a quicker and more accurate census. Automating census operations will allow us to replace labor-intensive and error-prone clerical operations with automated techniques that are more accurate and controllable.

Ms. **Oakar**. Doctor, thank you very much for your comprehensive statement. If you can, I would like to request that you wait until we have this other panel and then join them for questions. Is that possible?

Mr. **Keane**. I shall.

Ms. **Oakar**. Thank you very much.

Our next witnesses are Dr. Dan Horvitz, who is the executive vice president for Research Triangle Institute, which is a large nonprofit survey research organization based in North Carolina. Dr. Horvitz is also vice president of the American Statistical Association.

Prof. William Eddy is with the department of statistics of the Carnegie Mellon University, if he would come up, please. He has worked as a consultant for the Postal Service in the area of automation.

Ms. Judith Rowe is the associate director for research services of the Princeton University Computing Center. She is past president of the Association of Public Data Users and plays a prominent role among users in articulating the needs of university and private researchers.

And Dr. Benjamin F. King is the director of survey methods for the Educational Testing Service. Dr. King serves as the chairman of the American Statistical Association's Advisory Committee to the Census Bureau, and he has been a member of the panel on the 1990 census convened by the National Academy of Sciences.

We really do have a very distinguished group, and we would be happy to have your complete statements for our record. If you can summarize, it would be helpful so we could have more time for questions at the end.

So we will start with you, Dr. Horvitz.

STATEMENT OF DANIEL G. HORVITZ, EXECUTIVE VICE PRESIDENT, RESEARCH TRIANGLE INSTITUTE

Mr. **Horvitz**. Thank you, Madam Chairwoman.

I presume the statement will be included in the permanent record, and I will just make some comments.

Ms. **Oakar**. Absolutely. Your entire statement will be included for the record.

Mr. **Horvitz**. My remarks will be confined to the automated check-in system and the concurrent processing goal. Clearly the important goals of the 1990 census are to increase the efficiency of the census, the quality of the census and the timing of the census, and current plans certainly are moving in that direction through greater use of automation, through decentralization of the processing facilities and through concurrent processing.

In my opinion, the plans that the Bureau have been preparing have considerable potential for achieving greater accuracy. Certainly the automated check-in control, the automated control file, can help considerably to reduce the coverage errors that have occurred in previous censuses.

I was going to be talking about the potential for the computer edit and followup as a way of reducing content errors also over prior censuses. The remarks of Mr. Keane this morning suggest there is some reduction in the extent to which there will be computer edit followup as compared to previous censuses, and it is a little disappointing from my standpoint that there is going to be much more clerical edit, which was the process in prior censuses. Based upon his remarks this morning, this is anticipated in the information that was available to me with respect to the level of the current plans of the census.

Certainly from my standpoint, through automation, one has an opportunity to address problems with the census with respect to coverage and content early in the process, and all of my experiences show, it seems, that the closer you get to the source of the data in terms of time and space, the better job you will do. You will detect errors earlier and you correct those errors earlier. Automated processing helps this process considerably. So a reversion to a large segment of the population receiving only clerical edits I think puts us back where we were in 1980.

My written testimony certainly supports the plans and I certainly applaud the Bureau's efforts. I just think that the Bureau has been much too tentative in moving strongly in the direction of automation. I am surprised that we are here already in 1986 and not further along in the planning. Certainly it is possible that the Bureau, being challenged in 1980—with respect to the magnitude of the undercount in certain small areas and local areas—may have delayed excessively the effort available for the 1990 plans and that is unfortunate. The people who are the leaders at the Bureau with respect to planning, carrying out the census, and then answering the challenges are the same people, and they just may be overworked in that respect.

The cost of computer hardware and the complex software needed to more fully automate the census may be major stumbling blocks. Still, from my standpoint, the real potential for gains in quality and in productivity justify the added expense. My own experience in large-scale surveys has shown a very noticeable gain in both the quality and the productivity, despite the fact that getting set up for automation, it being a complex process, can provide many frustrations. But after you have things in place, properly tested and working, you forget those frustrations because of the gains that you see in the quality of the work that is being done.

Concurrent processing provides, I think, a major advantage in the ability to follow up and correct what would be called fail-edit cases in a timely manner. I feel that the census, particularly now in hearing what I heard this morning, is really not planning to exploit concurrent automation and the technology for concurrent automation that exists, sufficiently, particularly with respect to the followup procedure of fail-edit cases, certainly with respect to the quality of the edit process itself, if it remains a clerical operation for a good part of the census.

I have advocated that the technology exists for automating the process of followup in the telephone process by using computer-assisted telephone interviewing rather than paper and pencil telephone interviewing. Based on what I heard this morning from Mr. Keane, the telephone followup will be a paper and pencil operation. The computer-assisted telephone interviewing is a well-tested technology. It does have costs associated with the fact that each interviewer or enumerator working with a telephone must have a video terminal. The cost of video terminals are going down rapidly, and it seems to me that in the next few years the costs will be even less than they are today.

As I see the use of automation, we would have automated check-in, we would film, we will use FOSDIC, we will have computer readable data available concurrently with the collection of the data. That computer readable data can be transmitted back to processing offices from, say, any central office where the main computing is being done, can be transmitted via telecommunication back to the computers to support the computer-assisted telephone interviewing and processing offices. It does not appear that the Bureau is even considering the use of transmitting data via lease lines and establishing a network of their processing offices in which data are transmitted that way rather than shipping. The plan is to ship questionnaires. I do not think that that is necessary in this day and age.

I am also concerned that the technology exists and has been proven in the past to generate questionnaires for followup purposes as in the computer-assisted telephone interviewing. Basically the data would be in the computer. The enumerator would access that data in carrying out the interview, the followup interview to correct that fail-edit aspect. There are cases where you have to follow up in the field with a personal interview for households that do not have telephones, and in those cases also it is possible to generate out of the computer a questionnaire for the enumerator to use. The enumerator can be in one remote office from where the basic data are, and that data can print out that instrument for the enumerator to use in the local office.

Concurrent processing is, in my opinion, a very valuable approach, and the Bureau certainly is making headway in that direction. From my standpoint, it is not taking full advantage of the potential that exists with the technology that is available today.

I would like to just read a few things that I feel are also essential for the Bureau to consider from my written testimony.

I did not expect to see any discussion of Bureau plans to measure the quality of the 1990 census data at this stage, although I do know that they do have plans for the post-enumeration survey. I would like to suggest that there is room to consider components of error in a census in 1990 and would like to have the Bureau develop a plan for assessing the level of error introduced by all the check-in and data processing procedures used to establish the final data file record for each person enumerated and each household. This is a component of the total error. Clearly, there are many other sources of error, particularly with the quality of the coverage in the field and the quality of the data that is provided by the householders to begin with.

We have been conducting decennial censuses since 1790. Federal revenue sharing based on population and income data has enhanced the level of interest in census results considerably. The pressure to produce an accurate census, but within stringent budgetary constraints, is now heavier than ever. Yet the Bureau of the Census is expected to design and put in place procedures for collecting the data from about 100 million households in a short 2- to

3-month period, hiring and training a huge temporary staff to do so. From my perspective, the current time and budgetary constraints for conducting decennial censuses are inconsistent with the expected level of quality in small area data. There is a clear need to consider more rational census alternatives than our current decennial mobilization. It is also important to begin to consider such alternatives now if we are to be in a position to bring about a change prior to the year 2000.

One such alternative goes to the other extreme from current practice. It advocates collecting census data continuously in time and space covering approximately 1 percent of the population each month and approximately 10 percent each year. To explain, consider the 3,100 counties in the United States. It is possible to select 10 nonoverlapping samples of counties, such that each sample would have approximately 10 percent of the total population and also be representative of all the counties in the United States. Each sample will consist of about 310 counties. One of the 10 samples is selected at random each year without replacement and a census conducted in the counties included in the selected sample. In this manner, each county would have a census taken once every 10 years, which is also the case currently.

The quality of the census data with this alternative should be considerably greater than at present for several reasons, not the least of which is the amount of staffing required and the greater use of permanent staff. The country could be divided up into 31 regions with a permanent supervisory staff in each region. The staff in each region would be responsible for conducting a census in 10 counties each year, one county each month of the year, except, say, June and December. Thus the workload would be distributed uniformly over time and space.

An added advantage of this procedure is that accurate data on internal migration can be gathered, data which have never been available in intercensal years. Intercounty migration rates could be developed since each 310 county annual sample will produce data on in-migration for each census from all other U.S. counties and foreign countries, as well as data on out-migration from that county to 309 other U.S. counties. These current migration rates could be used to improve considerably intercensal year estimates of county populations over those computed currently. Thus, the allocation of Federal revenue sharing funds can be based on much better estimates of population in intercensal years than is possible with a census in all counties simultaneously only once every 10 years.

While I feel that costs should be less on some items and may go up on others with this process, the most important thing is the total cost would be distributed over each of the 10 years rather than incurring approximately 10 times the cost once every 10 years.

To summarize my remarks, I strongly support the current Bureau plans for the 1990 census, particularly the increased use of automation and the decentralization of processing facilities and the use of concurrent processing. I feel greater use of automation should be there. In particular, automation causes anybody engaged in operations as complex as the census to impose a much higher level of discipline to the data gathering and data processing enterprises, resulting, at a minimum, in greater staff productivity and, potentially, in a significant increase in quality. It is that disciplining, I think, that is an advantage in the end for large-scale undertakings such as this. That is really essential to produce the kind of census and the quality of census we desire. It is important to notice with automated procedures every case receives the same level of treatment in the census process.

In view of the complexity of the automated systems contemplated for the 1990 census, it is none too soon to begin implementation. I therefore urge the Bureau to come to closure on all aspects of the 1990 census processing procedures and to move rapidly toward implementation. This subcommittee and the Congress could assist the decision process at the Bureau by first demanding careful documentation of the computer hardware requirements to conduct an automated census at the planned level; second, based on this documentation, negotiating a mutually acceptable set of requirements; and, third, providing some assurance to the Bureau that the separate budget line item for the computer hardware included in the mutually acceptable set of requirements will remain intact. I recommend completion of this process no later than the end of this fiscal year.

Finally, I view the mode of conducting decennial censuses anomalous with the demand for even greater accuracy in the small area data reported. Therefore, I recommend that a serious effort be undertaken to examine alternatives to the decennial census for generating accurate small area data on the U.S. population.

Thank you.

Mr. ACKERMAN [presiding]. Thank you very much. I know that Chairman Garcia shares the same concerns as you do.

[The statement of Daniel G. Horvitz and his response to written questions follow:]

RESEARCH TRIANGLE INSTITUTE

STATEMENT
OF

DANIEL G. HORVITZ
EXECUTIVE VICE PRESIDENT
RESEARCH TRIANGLE INSTITUTE

TO THE
SUBCOMMITTEE ON CENSUS AND POPULATION
MAY 1, 1986

Mr. Chairman, I am Daniel Horvitz, Executive Vice-President of the Research Triangle Institute, a private, not-for profit research institute located in the Research Triangle Park of North Carolina. From its inception in 1958, the Institute has had a very active statistical research program, serving both the public and private sectors. It currently has a staff of 150 engaged in the design and conduct of social surveys relating primarily to health, education, and economic well-being. I am a statistician with over 35 years of experience in survey research. It has been my privilege to serve on the American Statistical Association Advisory Committee to the Bureau of the Census and I am pleased to have this opportunity to comment on the Bureau's plans for processing the 1990 Decennial Census. My remarks will be confined to the Bureau's technical papers covering automated check-in systems (Topic 2) and concurrent processing (Topic 3).

The primary challenges of the 1990 Decennial Census are to provide population data of higher quality in more timely fashion, and to do so in a more cost effective manner than has been achieved in the past. To meet these challenges the Bureau's current plans for processing procedures encompass several major innovations including greater use of automation, less centralization of processing facilities, and implementation of data processing in parallel with data collection.

The increased use of computer technology currently contemplated includes an automated address control file (ACF), automated check-in of questionnaires using bar code technology, automated generation and mailing of reminder post cards to nonrespondents, automated generation of nonresponse addresses for field follow-up, automated editing of completed questionnaires for inadmissible, inconsistent or missing items, and automated generation of fail-edit cases for either correction by an internal Edit Review Unit or correction by means of telephone follow-up (or personal visit follow-up, if necessary) with the household in question. The 1980 Decennial Census used manual procedures for all of these components of the collection process. Reminder post cards were not used in 1980.

Accompanying the contemplated increased use of automation in conducting the data collection phase of the 1990 Decennial Census is an examination of the potential for greater decentralization of the data processing functions. In 1980, all completed questionnaires were shipped to one of three processing centers where the data were converted to computer readable form. Cameras were used to take a picture of each questionnaire, the film was processed, and finally the images on the film were converted to computer readable data using FOSDIC (Film Optical Sensing Device for Input to Computers).

This same three-step process is contemplated for 1990 and is designated as FACT 90 by Census staff. The feasibility of increasing the number of FACT 90 locations from the three used in 1980 to as many as 24 in 1990 is being analyzed by Bureau staff. It should be noted that it is not necessary to confine all three FACT 90 steps to the same location. For example, cameras can be located at a large number of Processing Offices (PO's) to film the completed questionnaires. The undeveloped film can then be sent to FOSDIC centers for the remaining two FACT 90 steps. Bureau staff are also examining the feasibility of deploying cameras to as many as 50 locations with the film being developed and converted in no more than three locations.

Conversion of the questionnaire data to computer readable form in the 1980 Decennial Census did not begin until the data collection phase was complete. The plans for greater automation in 1990, including computer editing of questionnaires with telephone and personal visit follow-up of fail-edit cases that cannot be resolved otherwise, dictate the use of concurrent or parallel processing rather than the sequential procedure used in 1980. The decision to implement automated processing of completed questionnaires much earlier for the 1990 Decennial Census was made last fall at the Bureau's Decennial Census Decision Conference (DCDC). An extremely important consequence of this concurrent processing decision is the potentially significant improvement in the quality, timeliness, and cost effectiveness of the 1990 Census.

There is little doubt in my mind that the current plans for gathering and processing the 1990 Decennial Census have considerable potential for achieving greater accuracy than in the past. The use of automated systems for control of census operations involving mailings and field assignments, questionnaire check in, and telephone and field follow-up can reduce coverage errors significantly. These systems together with the planned computer edit for missing or inaccurate questionnaire entries should also result in considerable reduction, compared to prior censuses, in the number of clerical level staff required. The real beauty of these proposed systems is that they address actual errors occurring in both coverage and content early on in the process. They also have considerable potential for preventing errors that might otherwise occur. It has been my experience that better quality data are produced when systems are in place to detect and correct errors as close in time and space to their source as possible.

It should be recognized, however, that the proposed systems are complex. A number of major technical and logistical decisions remain to be made including the number and location of Processing Offices, the hardware and software required, the type and numbers of personnel to be recruited and trained, and the telecommunication network to be established. Currently, Bureau staff are addressing these issues through analysis of a series of action plans, using a set of common assumptions. In fact, the initial results of this analysis for the first 10 of these action plans were completed on March 14, 1986.

As indicated above, I heartily support the Bureau's automation and concurrent processing initiatives for the 1990 Census and congratulate their efforts. I have not identified any aspect with which I have major disagreement. On the other hand, it is quite possible that the Bureau has been too tentative about moving strongly in the direction of automation. I am surprised that it is already 1986 and that the Bureau is not further along in its planning than it appears to be at this moment. The demands on Bureau staff to address challenges and issues relating to the 1980 Census undercount may well have delayed earlier examination of the automation potential for 1990 at a sufficiently high intensity level. There may exist, among some Bureau staff, a feeling of satisfaction with the status quo, a strong inertial force often arising in the face of unwanted change, which may also have contributed to the slow rate of progress.

There is little doubt that the cost of computer hardware and the rather complex software development requirements can be major stumbling blocks to the adoption of a fully automated census in 1990. For these reasons I feel the cost and quality tradeoffs deserve to be developed and studied in some depth. While the hardware costs are far from trivial, particularly in the face of their relatively short period of use in a census, I am confident the gains in quality and productivity will justify the added expense. My own experiences with the adoption of automated procedures for use in the conduct of complex sample surveys have often produced frustrations initially, but then have resulted in such noticeable gains in quality and productivity that I now have difficulty remembering that there were any initial problems.

The development of an accurate, integrated system design for an automated 1990 Census with concurrent data processing as currently envisaged might best be accomplished using currently available computer-aided tools for systems development (such as Case 2000 or Excelerator). These tools use structured analysis and rigorous design techniques in a computer-aided environment that can lead to a significant increase in the quality of the detailed specifications, and hence in the overall quality of the ultimate system. I am personally not at all conversant with these tools, but mention them here since they might well provide the essential ingredients for addressing in an efficient and accurate manner the perceived complexities of a fully automated census.

A major advantage of concurrent data processing is the ability to follow-up and correct fail-edit cases in a timely manner. These are completed questionnaires which contain errors in the information provided for some items or for which the data are missing entirely on other items.

Current plans call for resolution of such fail-edit cases by an internal Edit Review Unit. In the event detected errors cannot be resolved by Census staff, the household will be contacted by telephone and new, presumably correct, data generated for the fail-edit items. Personal visits will be made to those households with unresolved fail-edits which do not have a telephone. Since the originally completed questionnaires will be checked in and stored in Processing Offices, this procedure presents logistical problems when a fail-edit case designated for personal visit follow-up is referred to a District Office (i.e. a data collection office) which is not also a Processing Office. Since there is no original questionnaire available in these District Offices (DO's) only a printed list of the items that failed-edit will be available to the interviewer. While Bureau staff recognize that new questionnaires containing the accepted original responses, but with the fail-edit items left blank, can be computer-generated for use in personal visit follow-ups in these latter DO's, it does not plan to do so for the 1990 Census, ostensibly because rapid full name capture is required, but not considered feasible.

In my opinion, there is much to be said in favor of generating new questionnaires, each with the accepted original responses but with the fail-edit items left blank, for use by the Census enumerators in personal visit follow-ups of fail-edit cases. The technology has been available for some time. My own experience, as director of the large National Medical Care Utilization and Expenditure Survey (NMCUES) in 1980, includes computer generation and printing, at remote sites, of two data collection instruments for use in subsequent interviews. The data generated for both instruments had been collected in earlier interviews. One of these instruments, called the Control Card, was a preprinted form, and contained all the demographic data about the family unit. Bar code identification consistent with the family unit identification number in the computer was also printed on the Control Card for automated check-in and control purposes.

The second instrument, called the Summary, was formatted entirely by the computer and contained data on all the medical services received and reported for each member of the household during prior interviews. There were fail-edit items in both forms. For example, the respondent may not have known the charge for a doctor's office visit reported in the previous interview because the bill had not yet been received at the time the respondent was interviewed. Therefore, the Summary for the current interview included all the information about the specific office visit and requested the interviewer to ask for the charge for that visit. The NMCUES experience in obtaining updated data on the Control Cards and Summaries was generally excellent. Essentially the same technology can be used in the 1990 Census, in my opinion, to generate questionnaires for field follow-up of fail-edit cases requiring a personal interview. Name capture seems unnecessary since the enumerator will have all the originally reported data for use in verifying that a particular family unit is the same one that sent in the original questionnaire.

Similarly, the Bureau staff recognizes that Computer Assisted Telephone Interviewing (CATI) can be used for telephone interview of fail-edit cases, but does not plan to do so for the 1990 Census. Without belaboring the point, CATI technology has also been around and proven for some time. It's value over paper and pencil telephone interviews is clear. The costs associated with providing the CATI interviewers with video terminals and the essential minicomputer support is most likely the primary reason for the Bureau's reluctance to use CATI for this purpose. A careful cost and quality comparison of the two telephone interview procedures for fail-edit cases before ruling CATI out would be very useful. Since there are now a rather large number of private survey and market research facilities with CATI capabilities, the Bureau might also give serious consideration to contracting with such facilities for telephone interview follow-up of fail edit cases. This would eliminate the need for a large investment by the Bureau in computer hardware to support the use of CATI for follow-up of fail-edit cases.

While I did not expect to see any discussion of Bureau plans to measure the quality of the 1990 Census data at this stage, I would like to suggest that the Bureau develop a plan for assessing the level of error introduced by all the check-in and data processing procedures used to establish the final data file record for each person enumerated and each household enumerated.

We have been conducting decennial censuses since 1790. Federal revenue sharing, based on population and income data, has enhanced the level of interest in census results considerably. The pressure to produce an accurate census, but within stringent budgetary constraints, is now heavier than ever. Yet the Bureau of the Census is expected to design and put in place procedures for collecting the data from 100 million households in a two month period, hiring and training a huge temporary staff to do so. From my perspective, the current time and budgetary constraints for conducting decennial censuses are inconsistent with the expected level of quality in small area data. There is a clear need to consider more rational census alternatives than our current decennial mobilization. It is also important to begin to consider such alternatives now, if we are to be in a position to bring about a change prior to the year 2000.

One such alternative goes to the other extreme from current practice. It advocates collecting census data continuously in time and space covering approximately one percent of the population each month and approximately 10 percent each year. To explain, consider the 3,100 counties in the United States. It is possible to select 10 non-overlapping samples of counties such that each sample will have approximately 10 percent of the total population and also be representative of all the counties in the United States. Each sample will consist of about 310 counties. One of the 10 samples is selected at random each year without replacement and a census conducted in the counties included in the selected sample. In this manner, each county would have a census taken once every 10 years, which is also the case currently.

The quality of the census data with this alternative should be considerably greater than at present for several reasons, not the least of which is greater use of permanent staff. For example, the country could be divided up into 31 regions with a permanent supervisory staff in each region. The staff in each region would be responsible for conducting a census in 10 counties each year, one county each month of the year, except, say, June and December. Thus, the census workload would be distributed uniformly over time and space.

This census procedure has an important added advantage. Accurate data on internal migration can be gathered, data which have never been available in intercensal years in the past. Intercounty migration rates can be developed since each 310 county annual sample will produce data on in-migration for each census county from all other U. S. counties (and foreign countries), as well as data on out-migration for each census county to 309 other U. S. counties. These current migration rates can be used to improve considerably intercensal year estimates of county populations over those computed currently. Thus, the allocation of Federal revenue sharing funds can be based on much better estimates of population in intercensal years than is possible with a census in all counties simultaneously only once every 10 years.

Costs for this census alternative should be less in view of a significant reduction in staff hiring and training costs compared to the current census procedure. Equipment costs, including computer hardware, should also be much less. There could be some cost increases associated with data processing and data dissemination. More importantly, the total cost would be distributed over each of the 10 years, rather than incurring approximately 10 times the annual cost once every 10 years.

Of course, one could dismiss this alternative quickly in view of constitutional requirements with respect to apportionment. On the other hand, if this suggested census alternative has sufficient merit on all the usual grounds by which census methods are evaluated, then a real potential for resolution of the apportionment issue exists, albeit not without considerable effort.

To summarize my remarks, I strongly support current Bureau plans for the 1990 Census, particularly the increased use of automation, the decentralization of processing facilities and the use of concurrent processing. The greater use of computer technology in the 1990 Census will impose an essential and significantly higher level of discipline to the data gathering and data processing enterprises, resulting, at minimum, in greater staff productivity and, potentially, in a significant increase in quality. It is important to recognize that, with automated procedures, every case receives the same level of treatment in the census process. For these reasons, I recommend the use of CATI for telephone follow-up of fail-edit cases and the use of computer generated questionnaires for those fail-edit cases requiring personal visit follow-up. In the face of strict limitations on the 1990 computer hardware budget, I recommend the use of contractors for CATI telephone follow-up of fail-edit cases. I also recommend the Bureau develop procedures for measuring the level of error introduced into the 1990 Census by the check-in and data processing phases.

In view of the complexity of the automated systems contemplated for the 1990 Census, it is none too soon to begin implementation. I therefore urge the Bureau to come to closure on all aspects of the 1990 Census processing procedures and to move rapidly toward implementation. This Subcommittee, and the Congress, could assist the decision process at the Bureau by, first, demanding careful documentation of the computer hardware requirements to conduct an automated census at the planned level; second, based on this documentation, negotiating a mutually acceptable set of requirements; and, third, providing some assurance to the Bureau that the separate budget line item for the computer hardware included in the mutually acceptable set of requirements will remain intact. I recommend completion of this process no later than the end of this fiscal year.

Finally, I view the mode of conducting decennial censuses anomalous with the demand for even greater accuracy in the small area data reported. Therefore, I recommend that a serious effort be undertaken to examine alternatives to the decennial census for generating accurate small area data on the U. S. population.

This concludes my remarks.

ANSWERS TO QUESTIONS ON 1990 CENSUS TESTIMONY

Prepared by Daniel G. Horvitz

Executive Vice-President
Research Triangle Institute

for the

SUBCOMMITTEE ON CENSUS AND POPULATION

of the

COMMITTEE ON POST OFFICE AND CIVIL SERVICE
U.S. HOUSE OF REPRESENTATIVES

Question 1. Could you tell us some more about how the Census Bureau could use computer-assisted telephone interviewing in the decennial?

-- How expensive do you think it would be to apply this technology to the 1990 Census?

-- Up to now has it been used for any operation as large as the decennial?

-- What leads you to believe that it would be practical for the census?

In my testimony, I recommended the use of Computer-Assisted Telephone Interviewing (CATI) for telephone follow-up of fail-edit cases and the use of computer generated questionnaires for those fail-edit cases requiring personal visit follow-up. In order to implement these recommendations, a specific level of automation must be planned and in place. Thus, the context of my CATI recommendation assumed a strong effort to minimize procedures which would require any further manual handling of census questionnaires, once they had been received and photographed in a Census district, regional or processing office. The level of automation contemplated by me for the use of CATI follow-up of fail-edit cases was consistent with, and among, the automation alternatives under review at the Bureau. Specifically, it included:

1. An automated address control file.
2. Automated check-in of questionnaires using bar code tecnnology.
3. Filming all completed questionnaires at the check-in office.
4. Shipping all film to FACT 90 locations for conversion to computer readable files and further computer processing.

5. A computer communication network linking the FACT 90 computer centers with the regional processing/field offices. The latter offices need sufficient computing power to support a CATI operation, as well as other data processing needs.

6. Automated editing of completed questionnaires at the FACT 90 computer centers and automated generation of a file of fail-edit cases for follow-up, either by CATI or by personal visit, with the household in question. The file should include all accepted original responses, but with the fail-edit items left blank.

7. Automated transfer of the fail-edit cases via the computer communication network to the regional processing/field offices.

8. Use of CATI in the regional processing offices to follow-up fail-edit cases that have telephones. Specifically, this implies an ability by the CATI interviewers to access, in turn, each of the unresolved fail-edit cases stored in the regional processing office computer.

9. Automated generation of questionnaires containing the reported data in field offices for field follow-up of fail-edit cases requiring a personal interview.

10. Automated transfer of the CATI resolved fail-edit cases via the computer communication network from the regional processing office to a FACT 90 computer center. Fail-edit cases resolved by personal interview, using a hard copy of the questionnaire, would be processed in the field office in the same manner as original hard copies.

The technology for the level of automation implied by the above exists. Subsets of the 10 components have been used in the past, either in prior censuses by the Bureau or in sample surveys by private sector survey research organizations. The major advantages to this level of automation are increased accuracy, timeliness and productivity, in my opinion. The major stumbling blocks are cost for the computer equipment, the ability to write and test the required software, and a willingness by the Bureau to commit to putting it in place for the 1990 Census no later than September 30, 1986.

The Bureau staff may feel the CATI portion will not be workable without having to key names of heads of households at check-in in order to be able to subsequently match follow-up cases with the correct household during the follow-up interview. The same potential matching problem exists for follow-up of fail-edit cases requiring a personal interview. This is not a problem for either the CATI follow-up or the personal interview follow-up cases, in my opinion. As outlined above, in items 6 and 9, the CATI interviewer, and the personal visit follow-up interviewer, will have all the originally reported data for use in verifying that a particular family unit is the same one that sent in the original questionnaire.

If the Bureau still has concerns about the ability to match fail-edit follow-up cases in an automated environment (i.e. without reference to the original hard copy), then I suggest the original questionnaire be structured so that the first initial and first four letters of the surname of the head of household can be coded by the household respondent for subsequent FOSDIC conversion to computer readable form. This additional information should be sufficient for linking CATI and personal interview follow-up cases to the correct original household without reference to the original hard copy questionnaire. I doubt that the Bureau has even considered requesting householders to special code surnames for subsequent FOSDIC conversion to computer readable form, yet current 1990 Census plans call for a significant amount of manual name keying.

I do not have access to the essential parameter values to know just how expensive using CATI for follow-up cases might be. I suspect the major additional cost item will be the computer hardware. There are cost savings since the follow-up data will not have to be keyed. There will also be savings associated with a shorter overall time period to complete the data collection. Since the hardware will only be needed for a short time, the Congress should authorize a separate hardware budget that would permit the Bureau to lease the hardware it requires.

Certainly, CATI has not been used for any operation as large as the decennial census. It has, however, been used in very sizeable operations, operations which have been as large as might be required for telephone follow-up of fail-edit cases by a census regional processing office. There is a tendency to discount the potential of census alternatives which have proven effective in sample survey operations by pointing out that the decennial census is very large. This is not a valid argument when, in fact, the same procedures are merely being duplicated across a large number of regional or district offices. The important question to ask is whether the technique or procedure is feasible and cost effective for a regional/district office. If it can be implemented for one office and work well, then it should be relatively easy to duplicate across the total set of census offices. For this reason, I don't consider the size of the census to be relevant to whether CATI or the other automated procedures listed above can be implemented or not in the 1990 Census.

I am convinced that the level of automation listed above is practical for the census, considering that only tested technology, including CATI, is suggested. The savings in manpower costs incurred with manual procedures may not be a sufficient tradeoff against the hardware costs incurred with the proposed automated procedures, however. It should be noted that hardware costs have been coming down fairly rapidly of late. It should also be recognized that automation could yield a significant increase in the quality of the census, so that any evaluation of the practicality of a particular automation alternative should be in terms of its cost effectiveness relative to manual procedures.

Finally, it is important, in evaluating automation alternatives, to compare total systems, in my opinion. It was not very evident from the background papers prepared by Bureau staff that total systems were being compared. I have a distinct impression that system components were examined piecemeal, and that they were evaluated only at the margin relative to 1980 Census procedures.

Question 2. How important is concurrent processing to the success of the 1990 Census? What will happen if the Bureau is not able to use this new system?

In my opinion, concurrent processing makes a great deal of sense. It can improve the timeliness of census reports considerably and at the same time reduce the cost of the census. I do not see how we can accept a lesser standard than that obtainable with concurrent processing. If the Bureau is not able to use concurrent processing, I can only conclude that the 1990 Census will be conducted no differently than the 1980 Census, and, as a consequence, with either less quality or, at best, with no change in quality.

Question 3. The background papers we received from the Census Bureau suggest that the Bureau might have to key the names of many people into its computers if it wants to complete the census on time and also do a full evaluation of its coverage.

— Do you think that this can or should be accomplished?

— Why is this needed?

— How would it affect the privacy of people who answer census questionnaires?

It is my understanding that the Bureau will key names to assist in subsequent linking of households in the post enumeration survey with households in the census. With more accurate matching, a better measure of the quality of the census will be obtained. There may also be other reasons associated with matching or linking of household records. As suggested above, if the Bureau considers keying names to be necessary then I don't understand why the Bureau has not come up with a better name capture technology, just as it came up with the FOSDIC technology to capture the census data from the census questionnaires some years ago. I consider the development of a system whereby census respondents would use a FOSDIC readable coding scheme on the questionnaire to enter the initials and four or five letters of the surname of the head of the household to be both feasible and practical. Thus, there would be no need to key the name data, since it would be read by FOSDIC equipment along with the factual data on the questionnaire. If name data for all persons in the household are required, then the FOSDIC readable coding scheme should be used for all. It should be noted that complete names need not be coded on the census questionnaire since the other census data known for a person can be treated as part of the signature of that person for matching purposes. My preference is for a scheme such as this for name capture rather than keying the names.

I doubt that keying the names would affect the privacy of census respondents. The Bureau has always put more than adequate procedures in place to protect privacy and confidentiality. I am confident the Bureau would do no less in this situation.

Mr. ACKERMAN. We will now hear from Prof. William Eddy, Department of Statistics, Carnegie Mellon University.

STATEMENT OF WILLIAM EDDY, DEPARTMENT OF STATISTICS, CARNEGIE MELLON UNIVERSITY

Mr. EDDY. Mr. Chairman, I am pleased to appear here today to comment on the preparatory work of the Census Bureau for the 1990 census. The remarks that I am going to make are excerpted from a written statement that I have submitted and I would like to have inserted in the record.

My comments are going to be divided into four broad parts: one part related to the automated address file, one part related to the automated check-in, a third part related to concurrent processing, and a final part which relates more generally to the use of computers and automated technology in the conduct of the census in the year 2000. It is not too soon to begin to think of that.

Although my comments are directed at technological issues, it is generally my impression that the most serious problems faced by the Census Bureau are managerial. The Census Bureau has demonstrated in the past its ability to solve these managerial problems, and therefore I think we need to challenge them to address these technological problems more directly. The potential benefits from the use of advanced technology in reduced costs, more accurate results and more timely results can be very large.

First with respect to the automated address file, every form distributed by the Census Bureau for the 1990 census will have a bar code label attached to it. Presumably the bar code labels will be generated by computer from the address control file. Thus, when a form returns, the bar code can be read automatically and the address control file can be updated to reflect the receipt of the form. This has such obvious benefits over the previous 1980 census procedures that the Census Bureau should be commended for its use. It will be much easier to maintain an up-to-date address list. When Bureau personnel learn of errors in the existing list, they can make changes quite quickly, easily, uniformly, permanently. This is just a vast improvement over the previous procedures which involved handwritten modifications to a printed listing.

On the other hand, there is an implied capability to communicate changes made to this file to other levels of the Census Bureau hierarchy and, I presume, to a central copy of this address control file. I am unable to discover that the Census Bureau has any plans to develop an automatic capability to allow this transfer of information.

With respect to the automated checkin, some of the documents that I was sent for review contained the following statement: "Would personal computers used in wanding"—that refers to read the bar code—"have more use after 1990 than laser sorters?" These are two alternative technologies. I think it is clear that personal computers are more likely to be useful than specialized laser sorting machines. In the 1986 test census the personal computers I think the question refers to were IBM PC/XT machines. The IBM PC technology was introduced in 1981 and by 1991 when these personal computers become available for other uses within the Census

Bureau they will be 10 years old. As is a problem with a lot of computer equipment, I think about amortizing my hardware over a much shorter period of time, say 4 years, because of the rapid pace of technology and, in fact, the IRS only requires that it be amortized over 5 years. The notion that IBM PC's can be used for any useful purpose by the Census Bureau in 1991 is absolutely ludicrous.

With respect to concurrent processing, I am surprised to discover that the Census Bureau has not done this processing in the past. From a management point of view it is obvious why they have not. On the other hand, the need for quality control seems to me to require processing of forms at essentially the same time as the forms are checked in, and I am very pleased to see the Bureau is moving in this direction and I would like to encourage them to continue more rapidly in that direction.

I would now like to turn to the future census, the 21st Century. We all know that the decennial census is the single most important statistical activity of the Federal Government, and our Census Bureau has a long and honored tradition of doing the best possible job for the lowest possible cost. Along the way the Census Bureau has been a harbinger of computer technology for the Federal Government and for the nation as a whole. However, based on my understanding of the planning for the 1990 census, this leadership role is changing. I fear that we may end up using 1970 technology during the census of the year 2000. I would, therefore, like to make a few suggestions which bear further exploration in an effort to move the census into the technological future. Of course, by the year 2000 it will be the present or even the past.

With respect to data collection, a review of the Census Bureau's planning documents that I have seen indicates no attempt to collect the raw data by the use of computer technology. Dr. Horvitz referred to a currently existing technology. I am thinking of something even more advanced. It is easy to argue that it is impossible to use computer technology in 100 million homes in the United States. I would challenge the Bureau to rethink this question.

The United States has the world's largest and most widely available data communications network, the telephone system. A large savings in manpower could be achieved if each respondent were to telephone the collection office and enter their responses directly into Census Bureau computers. The technology exists today to do this. Such a procedure does not even require that a respondent have a telephone. Merely access to a telephone. A further advantage, of course, is that respondents can initiate the call themselves rather than having some impersonal Bureau computer call them, and I presume that would improve the quality of the data.

With respect to data transfer, I was unable to find any hint that the Census Bureau has a plan to use modern technology for data transfer among its various offices such as computer communications networks and, in fact, in a specific documents referred to in my written testimony I found a flow chart adescribing the Census Bureau's plans for what they call basic urban processing, and in that document the symbol they used to indicate the transfer of data appears to me to be a railroad ore car of the type that were used in mines in this country in the last century.

Computer communications networks are ubiquitous in the United States and have been for a decade or more. They range from small local area networks which are typically privately owned and cover small geographic areas to wide area networks which are typically leased from the telephone company—I guess I should say companies—and are used for long haul data transmission over thousands of miles.

Two specific examples that were both developed and supported by Federal Government agencies are the Advanced Research Projects Agency [ARPA] net and the National Science Foundation NSF net. In a perfect system the Census Bureau microfilm system, the film optical sensing device for input to computers or FOSDIC, would be irrelevant; the data would be entered directly into a computer and could be stored on magnetic media or on laser disks. FOSDIC is one more example of the Bureau's predisposition toward tried and true technology. FOSDIC was originally developed in the mid 1950's for the 1960 census. In 1990 it is going to be refurbished and called FACT90. It will be 35 years old in 1990. In the year 2000 FOSDIC will be 45 years old. Surely the Census Bureau should lay plans to move from its paper-based methods and its paper-processsing techniques to something a little more modern. Sadly, I find no evidence to suggest this is true.

In particular, I believe that the Census Bureau must take a serious look at optical character recognition technology for use in future censuses. It is certainly too late to incorporate this technology in the 1990 census, but it is not too early to begin thinking and experimenting with its use for the year 2000.

The USPS has, I believe, extensive experience with OCR technology in its ZIP plus 4 program.

I would like to make the following final remark. The vast majority of the most serious problems faced by Census planners are the management of the estimated 100 million forms and the management of the people that are going to be needed to process them. Since the goal of the census is to determine or, to use a statistical term, to estimate the population of the United States as accurately as possible, it seems obvious that improvement to the management problem can be obtained by the use of additional approaches; that is, in addition to enumeration. The most obvious approach is to use a carefully designed sample together with sophisticated statistical techniques in addition to enumeration.

Since April 1980 there has been a lively discussion in academic, legal and political circles concerning the undercount revealed by the postcensus enumeration programs of the 1980 census. Much of this discussion has focused on whether we should enumerate the population or estimate it. I believe that estimation together with enumeration can be a highly cost effective solution to the management problems faced by the Census Bureau in 1990.

Thank you, Mr. Chairman.

Mr. ACKERMAN. Thank you very much for your testimony, Dr. Eddy, especially for a very graphic analogy.

[The statement of William F. Eddy follows; also included are his responses to written questions:]

May 1, 1986

STATEMENT
OF
WILLIAM F. EDDY
PROFESSOR OF STATISTICS AT CARNEGIE-MELLON UNIVERSITY
TO THE
SUBCOMMITTEE ON CENSUS AND POPULATION
OF THE
COMMITTEE ON POST OFFICE AND CIVIL SERVICE
U.S. HOUSE OF REPRESENTATIVES

Mr. Chairman, I am pleased that the Chairman and the Subcommittee maintain a continued interest in the preparatory work of the Census Bureau for the 1990 Decennial Census and I am pleased to have the opportunity to appear before you today to comment on that preparatory work. My comments are divided into four broad parts: the first part relates specifically to the use of an automated address file; the second part relates specifically to the automated check-in during the conduct of the 1990 Census; the third part relates specifically to the use of concurrent processing during the 1990 Census; and the fourth part relates more generally to the use of computers and advanced technology in the conduct of the Census in the year 2000.

I would like to make the following preliminary remark. Although my comments are directed at technological issues it is generally my impression that the most serious problems faced by Census planners for 1990 are managerial. Because the Census Bureau has demonstrated its ability to solve the managerial problems (in previous censuses) I think we need to challenge them to address the technological problems more directly. The potential benefits in reduced costs and more accurate results (and in less time) could be large.

1. Automated Address File

Every form distributed by the Census Bureau for the 1990 Census will have a bar code label attached to it; presumably, the bar code labels will be generated by computer from the Address Control File. Thus, when a form is returned the bar code can be read by automated equipment and the Address Control File can be updated to reflect the receipt of the form. This procedure has a number of obvious benefits over previous check-in techniques.

First, it will be easier for the Census Bureau to maintain an up-to-date Address Control File; that is, when Bureau personnel learn of errors in the existing file they can make changes quickly, easily, uniformly, and permanently. This is a vast improvement over previous procedures, which involved handwritten modifications to a printed list. On the other hand there is an implied capability to communicate the changes made to the local copy of the Address Control File to other levels of the Census Bureau hierrarchy and presumably to the central copy of the file; I do not believe that this capability currently exists or is contemplated.

The Address Control File is developed in cooperation with the Postal Service (USPS). I presume that the USPS has the ability to give a fairly precise list of the residents of any particular address and I am surprised that Census Bureau plans do not include the acquisition of this information. The potential savings in later keying of names are vast.

Second, check-in can be handled by automated equipment capable of reading the bar code label attached to the form, greatly reducing the clerical burden. The Bureau already has plans to use multi-pocket laser sorters and/or laser hand wands of the type that exist in many major retail stores today. I believe that the use of hand wands should be limited to forms that cannot be processed by the automated equipment but the Census Bureau plans are a step in the right direction.

2. Automated Check-in

One of the documents sent to me for review contained the following (I can only hope, rhetorical) question:

- *Would personal computers used in wanding have more use after 1990 than laser sorters?*

I think it is clear that personal computers are more likely to be useful after the Census than are laser sorters. In the 1986 Test Census the personal computers referred to in the question are IBM PC/XT machines. I feel compelled to point out that the IBM PC technology was introduced in 1981 and by 1991, when the personal computers are available for other use, will be ten years old. Most serious owners of computers amortize their hardware over a period of four years and even the Internal Revenue Service only requires it be amortized over five years. The notion that an IBM PC/XT will have any value to the Census Bureau in 1991 is ludicrous.

The Census Bureau has clearly decided to separate the processing associated with check-in from the processing associated with the gathering of information from the forms. From a management point of view this is clearly desirable. This is the stage where the confidentiality of the information on the forms is most weakly protected;

the actual check-in procedure is much simpler than the data-entry step. On the other hand, it appears that there will be some work which is repeated in the two stages, in particular, the keying of names appears to be duplicated.

3. Concurrent Processing

I am surprised to discover that, in the past, the Census Bureau has done no actual processing of the returned census forms until the check-in was complete. Again, from a management point of view this is the easy way. On the other hand the need for quality control seems to me to require processing of the forms at essentially the same time as the check-in. I am very pleased to see that the Bureau is moving towards concurrent processing for the 1990 Census and I can only encourage: more of the same.

I find it extremely difficult to understand why the Census Bureau finds it necessary to sort the paper forms. Surely it is more efficient to enter the data from the forms into a computer in the order the form are received and do the sorting with a computer. If it is essential to be able to recover a particular paper form, its position in the arrival order could be attached as a data item to the computer record for a particular form. I believe that the desire to sort the forms is just a bureaucratic habit and should be dispensed with; I challenge the Census Bureau to show that it is cost-effective to sort the forms.

4. Computers and the Census in the 21st Century

I know that we all agree that the Decennial Census is the single most important statistical activity of the Federal government and that our Census Bureau has a long and honored tradition of doing the best possible job for the lowest possible cost. Along the way the Census Bureau has been a harbinger of computer technology for the Federal government and the nation as a whole. Based on my understanding of the planning for the 1990 Census, this leadership role is changing. I fear that we may end up using 1970 technology in the Census for the year 2000. I would like to suggest a few ideas which bear further exploration in an effort to move the Census into the technological future (which will be the present or even the past by the year 2000).

The purpose of using computer technology in the Census is to speed up all the activities; as I am sure you know, even with the introduction of the Hollerith punched card machines, the 1890 Census still took nearly a decade to be completed. I am able to identify several major data processing activities (other than the actual manipulation of data that is traditionally called data processing) associated with the Census which can and should rely heavily on modern computer technology; these are:

1. data collection;
2. data transfer;
3. data storage and retrieval.

4.1. Data Collection

A review of the Census Bureau planning documents made available to me indicates absolutely no attempt to collect the raw data by use of computer technology. While it is easy to argue that it is impossible to make use of computer technology in the one hundred million homes that must be enumerated, I would challenge the Bureau to reconsider this matter for the years 2000 and beyond.

The United States has the world's largest and most widely available data communications network, the telephone system. A gigantic savings in manpower could be achieved if each respondent were to telephone the collection office and enter their responses directly into Census Bureau computers. The technology exists today: auto-answer modems, push-button telephones, pre-recorded messages, computer-controlled dialogs, etc. Such a procedure does not even require that a respondent have a telephone, only access to a telephone. A further advantage, from my point of view, to having respondents initiate the call is that the impersonal nature of human-computer interaction is reduced somewhat; of course, experiments with such a procedure might show that quality of the data is enhanced by having the Bureau initiate the calls with autodial equipment.

In what appears to me to be a very closely related matter, the government of France (which, I believe, owns the French telephone company, PTT) has had a plan to give away visual display terminals to every telephone subscriber. The improved accuracy of Census respondents who can visualize questions and answers to telephone queries seems obvious. This whole idea of having respondents respond by means other than pencil and paper requires that the Census Bureau relinquish the notion of having everything on paper or microfilm. Banks, which are unarguably at least as concerned with privacy and preservation of information as the Census Bureau, have been using computer technology for data entry and cash withdrawal for many years. Of course, the transition to a "cashless" society is taking much longer than its proponents thought; on the other hand, I, personally, have not entered a bank in the last several years.

4.2. Data Transfer

In my review of Census Bureau planning documents I was unable to find any hint of a plan to use modern technology for data transfer among the various offices. In the Processing Manual for the 1986 Test Census, (Volume IV, Chapter 1, Attachment A) even the Census Bureau itself has recognized how outdated its methods are; in the flow-chart describing the Basic Urban Processing, *the symbol used to indicate the transfer of data appears to be a railroad ore car of the type used in mines in the last century.*

I believe that the Census Bureau is planning to use magnetic tape recorded at 6250 bpi as its information transfer technology. This basic technology has existed for thirty or more years (with a density increase factor of twenty over that time period). This has the obvious benefits of being extremely reliable and familiar and the ovbious drawback of being extremely slow; data transfer from coast to coast takes days.

Computer communication networks are ubiquitous in the United States and have been for a decade or more. These range from local area networks which are typically privately owned and cover a small geographic area (a few square miles) to wide area networks which typically are leased from the phone company and are used for long-haul data transmission (over thousands of miles). I would like to mention two specific examples, both developed and supported by Federal government agencies.

About two decades ago the Advanced Research Projects Agency in the Department of Defense started development of a private packet-switching network for communication among its contractors and itself. Over time this network has developed into a world-wide data communications system with transmission times that are typically measured in hours. The network is composed of a large number of fairly short interconnected links typically operating at speeds of 56 Kb/second and is used for the transmission of large numbers of relatively small pieces of data. The stability of the ARPAnet as a communications system is legendary. The system is composed of redundant paths and sophisticated hardware and software; I believe there have only been two system-wide outages ever and the down-time during each outage was measured in hours. From the point of view of its users the ARPAnet is a tremedous success.

The National Science Foundation, as a part of its development of national supercomputer capability, is creating a high bandwidth packet-switching network to join its five supercomputer centers scattered across the country and to the large number of researchers who will need access to those centers. The network will be operational later this year. The important point to mention here is that there are a fairly small number of individual point-to-point links, connecting centers hundreds or thousands of miles apart, each capable of transmitting data at rates up to 1.5 Mb/second. According to a crude calculation, the entire Census data collection could be transmitted over a single link of this kind in a matter of months.

4.3. Data Storage and Retrieval

In a perfect system, the Census Bureau microfilm system (i.e., the Film Optical Sensing Device for Input to Computers or FOSDIC) would be irrelevant; the data would be entered directly into a computer and could be stored on magnetic media or on laser disks. FOSDIC is one more example of the Bureau's predisposition toward "tried-and-true" technology. FOSDIC was originally developed in the mid 1950's for the 1960 census. In 1990 it will be refurbished and called FACT90 ***and will be thirty-five years old***. In 2000, FOSDIC will be ***FORTY-FIVE YEARS OLD!*** Surely the Census Bureau must be laying plans to move from its paper-based methods and its paper-processing techniques to something a little more modern. Sadly, I find no evidence to suggest this is true.

It is a truism that personnel costs are increasing and equipment costs are decreasing. In slightly different terms, the cost of services is going up and the cost of goods is going down. If this is true, then any organization which engages in a people-intensive activity should be searching ***hard*** for ways to substitute goods for services. Computers and the associated technologies are an obvious place to begin. I believe that the Census Bureau must take a serious look at Optical Character Recognition (OCR) technology for use in future Censuses. It is certainly too late to incorporate this technology in the 1990 Census but it is not to early to begin thinking and experimenting with its use for the year 2000. The USPS has, I believe, extensive experience with OCR technology in its ZIP+4 (the nine digit zipcode) program.

In 1983, when I was consultant to a USPS vendor, OCR technology was already being used to assign nine digit zipcodes to individual pieces of mail and mark them with bar codes using inkjets. I believe that by now the USPS has purchased hundreds (or, perhaps, thousands) of machines which are capable of reading addresses and assigning nine digit zipcodes to the addresses for more than 80 percent of the mail pieces with less than two percent error; all this happens at rates like 25,000 pieces per hour per machine. My recollection is that these rates apply to "collection" mail (that is, the stuff they find in mailboxes) and that higher rates apply to various kinds of mass and machine-generated mailings. The remaining twenty percent that cannot be identified by the OCR equipment is set aside for human processing. As I recall the most difficult cases were addresses written with green felt-tip pen on green envelopes.

I would like to make the following final remark. The vast majority of the most serious problems faced by Census planners are the management of the estimated one hundred million forms to be processed (and the management of the people needed to process them). Since the goal of the Census is to determine (or, if I may use a statistical word, estimate) the population of the United States as accurately as possible, it seems obvious to me that an improvement to the management problem can be gained by another approach in addition to enumeration. The obvious approach is to use a carefully designed sample together with sophisticated statistical estimation techniques in addition to enumeration. Since April 1980, there has been a lively discussion in academic, legal and political circles concerning the "undercount" revealed by the various post-census enumeration programs of the 1980 Census. Much of this discussion has focussed on whether we should enumerate the population or estimate it. I believe that estimation together with enumeration can be a highly cost-effective solution to the management problems faced by the Census Bureau in 1990 and beyond.

RESPONSES BY WILLIAM F. EDDY TO QUESTIONS POSED BY CHAIRMAN GARCIA FOR THE RECORD OF THE HEARING OF MAY 1, 1986

1. **Question:** You have presented a very forward looking piece of testimony looking ahead to the year 2000. Considering how important advance planning is, that is certainly a laudable approach. But I wonder, how much of what you described is currently possible? What changes do you think that the Census Bureau could adopt for the 1990 Census that would implement your ideas with currently available technology?

Answer: The short answer is that all of my suggestions are currently possible; the technology exists today. Unfortunately, the management problems implied by attempts to implement the technology are myriad; and given the Census Bureau predisposition to resist changing a "working" system it may be impossible to implement any of my suggestions for 1990.

The long answer is that two of my specific suggestions realistically can and should be implemented for the 1990 Census:

1. the use of local computers rather than humans to sort questionnaires; and
2. the use of a communications network to transmit data to central computers.

The Processing Manual for the 1986 Test Census, Volume IV, Chapter 1 entitled "Urban Processing Overview" clearly indicates that the Census Bureau intends to sort the returned questionnaites into CO/CBNA/Block order following processing. *There is absolutely no reason to perform this hand sorting.* Since each questionnaire will have a unique laser-readable barcode attached to it, later searching, matching operations, or evaluations can easily recover a particular questionnaire from storage without prior sorting. This can be accomplished by numbering the returned questionnaires in order of arrival; while this numbering does not actually have to be marked on the returned questionnaires it probably should be. The numbering could be performed automatically by the multi-pocket laser reader/sorters by the addition to them of inkjet barcode writers. Having the number written as a barcode means that it could be subsequently read by the same laser sorter, if needed. If both the original barcode (which determines the physical location of the respondent) and the additional barcode (which determines the order of arrival and hence the ultimate storage location) are both entered into a computer (which, again, could be done automatically), then it will be possible to recover any particular questionnaire by merely performing a computer sort and select operation.

In my written statement I described, in some detail, two existing computer communication networks which were developed by Federal government agencies. I believe that the Census Bureau could achieve considerable speed-up in data transmission time with identical accuracy by the development and use of its own computer communication network. Given the organizational plans (a number of processing offices and a smaller number of district offices), it probably makes sense for the Bureau to plan on a hierarchical network corresponding precisely to the organizational structure; this is actually further justified by the data flow, which, for the questionnaires, will flow up the hierarchy; the point is that there will be little, if any, data flow between processing offices or between district offices. The largest number of communication links (those joining the processing offices to the district offices) do not need to have a large bandwidth and hence will have lower cost; the smallest number of links (those joining the district offices to Census Bureau central computer facilities) need the largest bandwidth but the smaller number of them will keep the cost down.

2. Question: In your testimony you criticized the Census Bureau's plans for the use of its equipment after the census. What do you think that the Census Bureau should do with its equipment after the census is completed? In an operation that will only last a few days, how can you justify spending money on expensive equipment that will not be of use after the operation is over?

Answer: The Census Bureau plans to use two general kinds of equipment in the 1990 Census:

1. highly specialized custom-built equipment with no other known uses; and
2. general purpose off-the-shelf equipment which is in wide-spread use.

The Census Bureau is planning to use a 24-pocket laser reader/sorter and a smaller 6-pocket laser reader/sorter during its processing of the returned questionnaires. This equipment will be built under a special contract to the Census Bureau and cannot be used for any other purpose. I find it very difficult to understand why the Census Bureau has not devoted a massive effort to find a way to avoid the use of such equipment for such a short period. It appears to me from the cost and processing-rate data supplied by the Census Bureau that it is not cost effective to use this custom-built equipment and it would be less expensive to hire and train individuals to use hand wands for the purpose of questionnaire check-in.

Off-the-shelf equipment which is in wide-spread use should *absolutely not* be purchased for short-term use. It is economical to rent/lease such equipment because there are others who can make use of it. Furthermore, the particular equipment that the Bureau apparently plans to use will be technologically obsolete by the time of its availability for other uses; I would hope that by 1991 the Census Bureau had moved its operations to more modern equipment.

Dr. Daniel Horvitz, in his written statement for the record of this hearing, has suggested that instead of conducting this massive once-every-ten-years effort the Census Bureau could conduct a continuing census which covered roughly one percent of the population each month. I heartly endore his suggestion, for it solves a myriad of dilemmas faced by the Census Bureau, not the least of which is what to do with specialized equipment. Under the continuing census there will be a continuing need for specialized equipment. This not only provides a better justification for building the equipment in the first place but also provides a longer period of use to amortize its cost. It is fairly obvious that there will be many other cost savings associated with a continuing census. Most of the activities associated with the decennial census will not be performed as one-shot activities with the attendant start-up costs but rather will be ongoing activities.

Mr. **Ackerman**. We will now hear from Dr. Judith Rowe, associate director for research services, Princeton University Computing Center.

Dr. Rowe.

STATEMENT OF JUDITH ROWE

Ms. **Rowe**. I have written testimony I would like to submit into the record.

Mr. **Ackerman**. Without objection, so ordered.

Ms. **Rowe**. Thank you.

I appreciate your kind introduction and I appreciate your invitation to provide testimony on the subject of the data products from the 1990 census and their dissemination. I think it is worth noting that among my credentials are several years as a representative to the Council of Professional Associations on Federal Statistics [COPAFS], including two as COPAFS chairman.

Since its early years, the Bureau of the Census has made data available to the public from each decennial census. Until 1960 all of these data were released in the form of printed reports. Although the formal 1960 publications program included only printed reports, early in the decade summary data for census tracts were made available on computer tapes for the 43 States which then contained census tracts. In the mid-1960's 1 in 1,000 and 1 in 10,000 public-use samples were released on an experimental basis in the form of 200,000, or approximately 100 boxes, of punched cards. In 1970 computer data products were incorporated into the regular publication program. These data were provided in two fashions: in the form of counts which corresponded to the 1980 summary tape files, and in the form of public use samples. The former, the counts, contained tables or aggregate data for areas as small as city blocks and as large as metropolitan areas, States and regions. The latter, the public-use samples, contained records written on computer tapes for 1 in every 100 households and for all of the people in those households, using 6 different samples.

In 1980, building on their 1970 experience, the Bureau of the Census rechristened their computer products, changing the counts to summary tape files and the public use samples to PUMS or public use microdata samples. In addition, data were written on tape in a more compact and efficient format; micrfiche products were produced from several of the summer tape files; a major printed report, the block statistics, was issued only in microfiche, supplemented by summar tape files; the quality of documentation for the computer files improved markedly; and the Bureau entered, and left, the software business.

During the decade since plans were made for 1980 census products there have been major technological changes in the computer world, namely the advent of microcomputers, and the prospects for additional changes are upon us. The Bureau's reputation as a publisher and distributor of data from the decennial censuses is impressive, both in the United States and abroad. However, in planning products from the 1990 census, the Bureau is faced with many unknowns which will affect the choice of the physical formats in which data will be delivered. The Bureau has already been involved for some time in an extensive program of public meetings from which they hope to learn about the product needs of their user community. Although this is a commendable project, there are three problems with it: one, the census using community grows from census to census, each time supplementing the experienced users with the novices; two, many, if not most, of the people participating in these meetings are recommending that the Bureau produce products appropriate for today's technology, not for the technology of 1992 and beyond; and three, many of the people who are most outspoken in describing their product needs have very narrow interests in terms either of subject matter, of geographic area or of geographic unit. I have heard a report of the Washington Census Products meeting and it seemed to have produced several sensible recommendations. One which I would particularly support is reducing from 100,000 the size of the areas which can be identified in the public use microdata samples or PUMS. In New Jersey, as in many other states, this would allow for the identification of all counties and would therefore reduce the need for detailed summary tables in either printed or machine-readable form.

When 1990 comes, we at Princeton, along with people at several other institutions will be providing access to our fourth decennial census in machine-readable form. There are others for whom this will be their first census. Although new users may lack a familiarity with basic census concepts, with the structure of census products, or with the ability to move around comfortably within the machine-readable products, careful thought should be given to how much training must be provided by the Bureau and how much can reasonably be provided by others. For example, if decennial census data in machine-readable form is distributed as part of the depository library program, who will train the depository librarians?

The technology for providing data to users is changing rapidly. Although many users today are requesting that the Bureau make data available in the form of floppy diskettes, it is already apparent that this is an inefficient way to make large amounts of data available. For example, a recent request for Summary Tape File 1, the smallest STF, for a single country in New Jersey, Morris, required 32 diskettes. It seems more likely that by 1992 it will be more appropriate, efficient and inexpensive to download from online services such as CENDATA or from local services provided by State data centers or special services provided by commercial vendors. Alternatively, users who are currently using microfiche or having tables printed out from summary tapes may find access to CD-ROM the most suitable solution.

The Bureau has had many years of experience in melding together the diverse needs of users. I trust they will be able to compromise these needs and that those who advise them will not be intimidated by those who yell the loudest.

Some decisions, however, are affected by other considerations and it is in these areas that the Bureau has already made some sound recommendations. these are likely to lead to the early release of data. For example, by omitting from the first reports data from statistical areas which are created on the basis of the data collected by the census and for multi-State areas, it will be possible to make data for governmental units available sooner. Delaying the publication of historical data will also contribute to this goal.

Not only do users want particular subject matter, for particular areas in appropriate formats delivered promptly, but they want that data to be correct. It is a monumental task to insure that all of the census products are released without error, but this should be a primary goal. If there are errors in the census data, no one else can correct them. The Bureau of the Census must focus on those tasks which only they can do and leave tasks which can be done equally well to the State Data Centers and to the commercial sector.

For example, except perhaps for a simple retrieval package, the Bureau should not write software. By producing generalized data dictionaries, the Bureau can save other software developers much time and effort, and provide equally for all of them.

The Bureau should produce general purpose data products in a variety of formats which can be used by others to produce special-purpose products to meet individual needs. For example, the Bureau should not produce custom diskettes from summary tapes anymore than they have produced tape extracts from summary tapes.

I look forward eagerly to the 1990 census and to the products which will be produced from it. It is exciting to see the growing use of decennial census data and the varied purposes for which it is used. If the taking of the census is a national ceremony, surely the data which result are a national treasure.

Thank you.

Mr. ACKERMAN. Thank you, Doctor, for your testimony and your complete testimony is entered into the record.

[The statement of Judith S. Rowe and her response to written questions follow:]

STATEMENT OF JUDITH S. ROWE, PRINCETON UNIVERSITY COMPUTING CENTER

INTRODUCTION

We take for granted the quality, the quantity and the ease with which we access decennial census data. The data products created by the United States Bureau of the Census have become a model for other federal statistical agencies as well as for statistical agencies throughout the world. But even models sometimes have flaws.

In the past decade the data user community has grown in both size and diversity. Although we would all subscribe to a recommendation that data be accurate, prompt, useful and usable, we do not all mean the same things by such a recommendation. As Ed Pryor of Statistics Canada noted in addressing the Second Census Research Conference, "We cannot assume that everyone who uses data advances at the same pace." Pryor referred to the "fast data approach," comparing it to the fast food approach, and contrasted it to the indepth needs of research demographers.

We must also contrast the substantive data needs of users in different parts of the country. For example, I can live with data errors in Wyoming, but data users in Wyoming cannot. I can live without detailed ethnic breakdowns for census tracts, but the city of New York cannot. I can live without detailed data on American Indians but the Bureau of Indian Affairs cannot. The decisions concerning products from the 1990 census, which will be addressed first, focus more on format than on content. Specific content decisions will come later.

Although neither format or content as such, it is important to note that the Bureau must invest more effort than it did in 1980 in insuring that the data are accurate before they are released. It is true that everyone wants data to be released promptly but not at the expense of accuracy. It is important to avoid the public proliferation of tapes with errors, as in the case of Summary Tape File 2, for which a correction tape was issued, or Summary Tape File 3A, for which replacement tapes were issued. We have no way of knowing how many purchasers actually corrected their STF 2 files or replaced their STF 3A files and, in any case, I would guess that the efforts to correct these errors were more costly, both to the Bureau and to the user community, than extra pre-release verification would have been.

The past decade has witnessed enormous technological changes in storage media and in computing devices and there are, therefore, far more options for producing data products than ever before. It was easier to make recommendations in 1976 and it is likely that there will be sufficient stabilization in computer storage technology to make it easier again in 1996.

I remember serving on a committee which was making data product recommendations for the 1980 census. Although we recommended some modifications in the 1970 products we did not recommend new formats for providing data. In 1982 our recommendations still seemed reasonable. I hope our 1986 recommendations will fare equally well. In order for this to happen we must anticipate today the formats which will be commonplace in 1992.

The Bureau is now circulating a paper titled "1990 Census Products Issues" which provides general guidelines and a large range of specific questions to which they are seeking answers. The paper is based on discussion within the Bureau and at numerous public meetings held within the past two years, as well as on the advice of advisory committees and of individual users and information intermediaries such as librarians and State Data Center (SDC) staff. It has been no mean feat to mediate the recommendations of so many diverse users, users with diverse informational needs and technological capabilities.

Following a general planning section the paper raises questions in four major subject areas. These are:

1) statistical reports, microfiche and tabulation contents;
2) computer tapes, including summary tapes, reapportionment/redistricting data, public-use microdata files, data for microcomputers and in other formats, documentation, special tabulations, and software;
3) maps and geographic reference products, including the Master Area Reference File, census tract and block maps, geographic-area equivalency files, and area measurements; and
4) services to data users, including communications, guides, and indexes.

By and large my comments will be directed to the Bureau's "issue" paper. The focus of the paper purports to be the 100% data, the short form questionnaire which must be completed in every household, but much of the paper applies equally to the sample data. A subsequent paper will specifically address the sample data, the full-range of housing and socioeconomic questions which are collected only from approximately 20% of the households in the country using the long-form questionnaire.

I. PLANNING FOR 1990 CENSUS PRODUCTS AND SERVICES

The Bureau of the Census' issue paper on 1990 data products begins with some general guidelines and with a schedule for data delivery. The latter calls for the publication of all data products by the middle of 1993. This is an acceptable schedule; the question is "what can the Bureau do to adhere to it?" The original schedule for 1980 was also an acceptable one, but the Bureau fell far short of it.

Perhaps the most specific and the most imaginative of the proposals the Bureau is offering for meeting its schedule is that data for governmental units be made available without waiting for data for statistical areas. Summaries for metropolitan statistical areas (MSA's), urbanized areas or census-designated places are based on the data collected by the census and cannot be prepared until all of the data for each area's component parts are tabulated. Further delays will be avoided by postponing publication of data for multi-state areas such as MSA's, urbanized areas or Indian Reservations. Other commendable ideas for speeding up census product production are the separation of historic data into separate products and the reduction of the amount of race and ethnic data in the standard printed reports.

There are some plans under consideration for improving fiche products and for developing products for the novice user. Given the cost and the proposed attention to new media, I would question the value of too much investment in improving the quality of the microfiche products. I also fail to understand a recommendation, made in another Bureau paper, to produce on-demand printed copy from microfiche. Surely this is a function of the State Data Centers or of the commercial sector. The same goes for products like video tape, graphic summaries, or on-demand pamphlets with extracts from summary tape products, all proposed as a means of serving novice users. Whatever effort is expended on the indices for the summary tape products can and should be extended to the microfiche. For limited use the Bureau-produced fiche is satisfactory and increasing the ease of using machine-readable products such as CD-ROM, for example, should preclude the need for microfiche in the future.

II. PRINTED REPORTS AND MICROFICHE

A. POPULATION AND HOUSING COUNTS FOR GOVERNMENTAL AREAS

As indicated above, the need for prompt data for governmental areas is extremely great. Since MSA's are composed of counties, users can at least reconstruct them as they previously existed. The general principle of releasing available data as soon as possible is excellent and providing 100% data for governmental areas without waiting for sample data also makes good sense.

Users often photocopy pages from printed reports. This would be expedited by improving the quality of the binding, by using gathered, sewed or stapled bindings instead of perfect bindings. The use of margins would also improve the quality of copies. Many of the printed reports from decennial censuses are heavily used and, although hard covers would be ideal, better paper covers would be some help.

B. BLOCK STATISTICS REPORTS

The experience of 1980 indicates that there is no necessity for publishing printed block statistics reports. Since blocks will now be available nationwide, it seems reasonable to produce them by state or county group within state rather than by metropolitan area. Although it may be premature to make this decision, CD-ROM may well be a felicitous substitute for 300,000 frames of fiche. Block data are typically used for the study of local areas, seldom for the country as a whole. Therefore, for most users the availability of more blocks will not justify any but the most trivial loss of subject material.

The block statistics maps are the basic census maps. It is essential that they be available to users in hard copy and there seems no reasonable way of doing this except by printing them. However, it might be well to substitute the smaller size of the 1970 block maps for the larger and somewhat more awkward size of the 1980 maps.

C. COMBINED POPULATION AND HOUSING REPORTS

It is acceptable to combine 100% housing and population data in the printed reports, as has previously been done with the summary tapes, and to publish these data prior to the availability of either the sample data or the statistical area data. However, it is not acceptable to omit data for MCDs in 11 states, particularly as in New Jersey, where all of those MCDs are governmental units. We did battle on this issue after the 1970 census and it was my understanding that it had been resolved. I have never been able to explain to users why there was printed information for Princeton Boro, which was a place, but not for Princeton Township nor why Woodbridge Township and Hamilton Township, neither of them places but both with populations of over 100,000 did not appear in the 1970 printed reports. I am not comfortable with the omission of data for smaller places and I have never been able to understand the logic of printing data by population size groups. If you don't already know the size of the area, you don't know where to look for information about it. It is not necessarily true that the amount of information one needs about an area is a function of its size.

D. REPORTS COMBINING 100-PERCENT AND SAMPLE DATA

It does not seem necessary to separate the publication of the 100% census tract data from the publication of the sample census tract data in the way that is being proposed for larger areas. However, information about the availability of the 100% data in non-print formats should be made more widely available than was the case in 1980. In 1970 Newark Public Library produced a list showing all of the 567 municipalities in New Jersey and the tracts which each contained. If we had a comparable list for every state in 1990, it wouldn't make any difference in what order the tract data were published. Numeric order by county would be fine. The proposal to provide only metropolitan tracts in printed form seems acceptable. It would seem reasonable to consider much of the race and Spanish-origin detail published for tracts in 1980 as candidates for special tabulations. Because of the changes in tract boundaries, the creation of new tracts, etc. I would not recommend providing 1980 population and housing counts; in many cases this information will be misleading. I would appreciate having land areas and, armed with this and tract population, I think most users could compute density by themselves.

E. HISTORICAL COUNTS AND DATA

I see no need to publish historical data for either metropolitan areas or for tracts. It is certainly useful for governmental areas, in both the printed reports and on the summary tapes, but no reports or tapes should be delayed in order to include these numbers. They can always be obtained from the publications of earlier censuses.

F. MICROFICHE

Without a good sense of the amount of use the 1980 fiche received, and without a sense of the costs involved in improving the quality of the fiche, it is difficult to comment on the proposal to use a more expensive fiche product, one which would generate a substantially greater number of frames. It is not hard to read the fiche currently in use on the screen and most users do not require hard copy. Reader-printers rarely produce easy-to-read copies, even of the best fiche, and it seems foolish to expend too much money or effort on a transient technology. As indicated earlier, the one effort which seems warranted is one which would provide, if not an index to the fiche, at least some meaningful labeling on the fiche themselves. If the cost is not prohibitive, I would recommend doing both.

G. RACE AND SPANISH ORIGIN DATA

It seems appropriate to reduce the amount of subject detail cross-tabulated by race and Spanish origin in the regular printed reports. The real question will be which tables to eliminate. If it is necessary to decrease the number of areas for which race and Spanish origin data are produced, raising the required population of a group in order for an area to qualify for inclusion (e.g. from 400 to a higher number) would seem easier for library patrons than switching to a percentage system. This is not a burning issue; using the Trenton SMSA as an example, raising the minimum from 400 to 800 would probably reduce the relevant section of a report now containing 46 pages plus 47 pages of appendices from 18 to 12. What is this saving in dollars and cents? Could this money be better spent on special subject reports?

III. MACHINE-READABLE PRODUCTS

A. SUMMARY TAPE FILES

The first machine-readable product is the reapportionment/redistricting file mandated by Public Law 94-171. The reapportionment counts are scheduled for release by December of 1990; the PL 94-171 file by March of 1991. A survey of the state officials who used these data for the 1980 reapportionment and redistricting reveals that they were generally satisfied and that the few problems that they did identify seem likely to be addressed by the Bureau's plans for 1990, the complete blocking of the entire country and the development of the TIGER system to improve the quality and consistency of the census geography and maps.

As with the printed reports, the Bureau is attempting to make 100% data available earlier by providing it without waiting for the statistical areas, without the addition of historical data and without the inclusion of areas which cross state boundaries. This is a commendable idea since each Summary Tape File (STF) will eventually have a national level file which will include these areas. If the Bureau plans to produce a printed report with historical data I fail to understand why producing a machine-readable one would be any more effort. A congressional district file for the 102nd Congress would seem to be a useful reference since in some states districts will change quite dramatically after reapportionment. However, my own recollection is that there was actually little call for these data.

The enormous increase in the number of blocks nationwide would certainly argue for separating the blocks from the other geographic areas. Using the same table format for both would provide the advantages of 1970 when blocks were separated and the advantages of 1980 when they were merged with the larger areas.

The proposal to provide two versions of STF2 and of STF4, one for total population only, seems quite sensible. The majority of users of these files used only the total data. In fact much of our use of STF2 would have been eliminated if single years of age had been included in STF1.

One 1980 initiative which should be continued is the use of area names. Although it is useful to have some area identifier on each segment of a logical record it is only necessary for the first segment to have as long an identifier as was used in 1980. My guess is that most people use FIPS codes and that the census codes could be omitted. Among the other useful additions in 1980 were the complete-count variables on the sample files and the addition of medians. Another useful addition would be median age. It is my impression that we had more requests from the 1980 census for data cross-tabulated by age than was the case from earlier censuses. This was over and above the new tabulations for people 60 or 65 years and over.

There is no question that the use of split tracts caused numerous problems. The most obvious, but perhaps the most difficult, solution to this would be to get rid of them. Including summaries for tracts, as well as for BNA's or places, that are split by a higher geography would certainly help but it is difficult to recommend a subject matter deletion which could be made as a trade.

Although I am sure that there would be some use of an STF 1 zip code file it does not seem to be a high priority item. The creation of SPSS "export" files might well antagonize two sections of the private sector: the producers of value-added census files and the other statistical package vendors. I recognize the Bureau's desire to ensure that an easy-to-use product is available to the public, but I doubt that this is the answer. Certainly it is not a substitute for the more generalized format know in use.

B. PUBLIC-USE MICRODATA FILES (PUMS)

The 5% sample was unquestionably the most heavily used, both because of its size and because of its structure. The metropolitan sample received somewhat less use, although it is possible that had it also been a 5% sample it would have been more popular. The urban-rural sample was probably almost as unpopular as the 1970 Neighborhood Characteristics Sample. I cannot explain the unpopularity of either. There is no doubt a market for separate microdata samples for age, employment status and a variety of other characteristics as well as race and ethnicity. However, it seems more appropriate for the Bureau to draw the line at producing general purpose samples and then allowing others to select special populations from them. Although it does not seem necessary to add family records to the PUMS, family sequence numbers on the person records would make intra-family as well as intra-household analysis possible.

C. DATA FOR MICROCOMPUTERS

Microcomputer technology is changing almost daily. It is almost impossible to decide today how best to provide data for access by microcomputers in 1992. However, if I can blue-sky a bit, I would anticipate that in 1992 floppies or flexible diskettes will have gone the way of the punch card; that laser disk products will be more standard, less expensive and therefore more available; that the typical hard disk will be many times larger than is the case today and that micro-mainframe communication will be ubiquitous. I think this is a conservative prophesy, one which we can plan for with almost absolute certainty. Being a bit less conservative, I would guess that there are likely to be far more advanced laser products available and that micro-mainframe communication will be substantially cheaper. Under any circumstances I would assume that in 1992 CD-ROM will be less expensive to produce and that many, if not most, users will have access to the necessary technology. In terms of function I see CD-ROM as a substitute for microfiche and as the source for on-demand publishing. Therefore, it is entirely appropriate for the Bureau to confine itself to the development of simple display and extraction software and to leave the development of analytic tools to the commercial sector. I do not yet see a practical substitute for magnetic tape; rather I see it as a continuing source of extracts to be downloaded and as the medium for large-scale analysis. Although I do not anticipate a long-term need for either the Bureau or other data distributors producing custom diskettes, there is likely to be a period within the next three to five years during which the two technologies are likely to overlap. Unless I am very mistaken this period should be over by the time the Bureau is ready to release 1990 census data.

D. DOCUMENTATION

The quality of the documentation that has been released with the decennial census data is excellent. Although the User Note system is awkward unless the documentation is in a loose-leaf format---and that presents the missing page problem---there seems no better solution. My only recommendation is that each edition of the documentation include an edition number and/or date and the numbers of the User Notes it incorporates. If it did not add excessively to publication costs, I would argue for publishing the documentation in BOTH bound and loose-leaf form.

The Bureau should develop data dictionaries in a "generic" format that could be converted by users for use with any software. This seems more efficient and less costly than trying to develop separate dictionaries for each of the software products used with census data. The dictionaries should be either sold separately or made available by subscription and released on a flow basis. Many users will want to start work with both the documentation and the machine-readable dictionaries as soon as they are available;

others will have no interest in the accompanying material until the data are available and still others will acquire their dictionaries from some other source.

E. SPECIAL TABULATIONS

My understanding is that the Neighborhood Statistics, the EEO (affirmative action) and the School District files were all well-used. If this be the case, I suggest they be continued. As for other types of areas, defined by users in terms of census blocks, this is only sensible if it can be done efficiently and at a reasonable cost. If standard tables are used this should be the case; if users wish to produce custom tables for custom areas, the cost to the user should increase substantially.

The idea of producing "subject files" with data for counties in much the same way as in previous years "subject reports" were produced for regions or states is a good one. It is my belief that the subject reports were sorely missed and that having data for counties would greatly enhance their usefulness. These subject files seem a good candidate for a CD-ROM format.

F. SOFTWARE

The Bureau should do what the Bureau can do best, and writing software is not it. As far as I am concerned CENSPAC was an expensive fiasco which I would not like to see repeated. I would assume that the software which the Bureau develops for internal use will have certain special features which would make it inappropriate for general users.

G. OTHER

I realize that in order to protect the confidentiality of census respondents it is necessary to suppress or modify some of the published data. I know that there were complaints about the flag system which was used to indicate suppression on the 1980 summary tape files. However, I find the two other solutions of which I am aware less satisfactory than the one in current use. The system used in 1970 was awkward and confusing and the random rounding system being considered requires constant explanation for tables whose cells do not add up correctly. Unfortunately most data users would rather that numbers add up, even if they add up to the wrong total.

IV. MAPS AND GEOGRAPHIC REFERENCE PRODUCTS

Although the development of the TIGER file will not radically change the needs of the user community for maps and other geographic reference products, it should radically change the Bureau's ability to produce these products. Unfortunately the Bureau's census products issue paper does not address these concerns. Obviously there are some products which should be produced at any cost and other products for which even a small expenditure would be unjustified. I am not a map expert and therefore I can only share with you my own satisfactions and my own annoyances in using the Bureau's geographic products.

I cannot at this point comment on the proposed TIGER products. As for the Block Numbered Map Series, my primary concern is that there be block maps available in paper form, of a size similar to the 1970 maps, and that preferably they be organized by counties. Although the addition of a second color to these maps is desirable, it is not essential. Additional map features are always a plus; they help you locate blocks in unfamiliar areas. However, some reasonable compromise is certainly acceptable. I made little use of the Reference Maps in the past, but I may not be typical. The thematic maps were very attractive but I am not sure that their production is really a Bureau function. Armed with the appropriate data, any of the many academic or for-profit organizations with mapping capabilities could produce similar products, perhaps even at moderate cost. The ability to do this would, of course, be enhanced by the availability of Digital Geographic Area Boundary Files. The complete digital files are very large and the Bureau should make them available on tape. It might be reasonable to make "thinned" state and county files available on diskettes, although they are widely available commercially. As for producing "thinned" files for smaller substate areas, this seems a reasonable function for the State Data Centers.

The Bureau should continue to produce the Master Area Reference File (MARF). It is very convenient to be able to get all of the small-area geography for the whole country on a single tape. However adding some key variables such as land area to the Summary Tape Files will obviate the necessity for producing more than one edition of MARF. This will be a saving for both the Bureau and for the data centers which acquire its products. I don't feel that I have adequate information to comment on the Address Reference or Geocoding Files.

My experiences is that both of the Geographic Reference Products were well-used. The Geographic Identification Code Scheme (GICS) is a quick reference for locating MCDs/CCDs, and places, and for finding appropriate codes for them, and for states, counties and metropolitan areas. Although Equivalency Files between censuses have some value, it is difficult to explain the nature of boundary changes in tabular form. The same is true for equivalencies within censuses as in the cases of school districts, neighborhoods and other similar districts. I would welcome an imaginative way of dealing with these problems. On the other hand, it is enormously useful to have in summary form the places in a given county group area or the tracts in a given county.

V. SERVICES TO DATA USERS

A. STATE DATA CENTER PROGRAM

Princeton and Rutgers Universities have been providing access to machine-readable census data since the early 1970's and, through the aegis of the Joint Committee on Printing's Depository Library Program, to printed census material since the 1860's and even earlier. The JCP is exploring the addition of data in electronic format to the Depository Library Program. It is likely that by the time data are released from the 1990 census not only Princeton and Rutgers but all state and major public, academic and law libraries will be developing the facility for making machine-readable data available to their patrons, probably via microcomputers. It behooves the State Data Centers to prepare for a greater involvement of libraries in the area of data delivery. I am pleased to see that libraries are already being involved in State Data Center meetings. In addition, however, the major libraries, many of whom are already actively involved in the SDC program should think more seriously about how to serve the small public and school libraries better.

B. COMMUNICATIONS

All of the Bureau of the Census' current newsletters and fliers have well-defined purposes and well-defined audiences. There seems no reason to produce an additional newsletter. *Data User News* is the general-purpose, large audience newsletter which reaches individuals interested in any or all of the Bureau's products. *Monthly Product Announcement* provides quick information on new Census products of all types, those already released and those soon to be announced. *Fact Finders for the Nation* are excellent brief reference tools which describe all of the sources of information, regardless of format, on a given subject and are brief and inexpensive enough to be acquired in quantity. The chart on pages 6 and 7 in the revision of the *Factfinder for the Nation* on "Data for Small Communities." is an excellent model for future *Factfinder* products. *Data Developments* provides a data service or library with abstracts for all of the Bureau's machine-readable products, as they are released. This makes it unnecessary to acquire the full technical documentation or codebook unless the data themselves are being acquired. CENDATA can be used as a means of providing prompter information about new products. Handouts describing each of these information sources as well as the *Users' Guide* and the Indexes could provide guidance in their use. This would increase their use, create more informed users, and fewer inquiries to Census offices.

C. 1990 CENSUS USERS' GUIDE AND INDEXES

A *Census Users' Guide* is essential for users of census products. However, it is important that as much of the *Guide* as possible be made available in 1990 along with skeletons of the data products, a glossary and indexes. I would prefer having a subscription to the *Guide* and receiving replacement pages as they are issued. However, I recognize that for libraries and other places at which they provide more public access to their publications this is difficult. Could we have it both ways?

D. CENDATA

CENDATA is developing a growing audience, one which will increase as it becomes available through less expensive services. It can be used effectively to provide new data for large areas as they become available, new product announcements and, if possible, as an ordering service. CENDATA offers a useful means for providing on-line access to national data, data for states, metropolitan areas and large cities. I question its cost-effectiveness for smaller areas.

CODA

We stand at an exciting time as we view the prospect of a growing body of users of federal information, a substantial portion of whom are users of data from the decennial census. In an age of technological change we see the role of the computer as a means of producing these products as well as a tool for using many of them. The Bureau of the Census' efforts to poll the public for their advice in the design of the products which will emanate from the 1990 census is commendable. We've come a long way from the single volume which summarized the early censuses.

Princeton University
Computing Center
87 Prospect Avenue
Princeton, NJ 08544

May 8, 1986

Congressman Robert Garcia
U.S. House of Representatives
Committee on Post Office and Civil Service
Subcommittee on Census and Population
219 Cannon House Office Building
Washington, DC 20515

Dear Congressman Garcia,

I welcome the opportunity of responding to your thoughtful questions. I am sorry that you were not able to be at the hearing and I look forward to meeting with you at another time. However, I did appreciate the opportunity of appearing before Congresswoman Oakar and Congressman Ackerman.

I am enclosing a clean copy of my oral testimony. The answers to your questions appear below.

1. I understand your concern about my advocating the use of CD-ROM as a medium for providing access to the 1990 census data and I share your concern for end users. However, it is my belief that the State Data Center program should serve as the intermediary between the Bureau of the Census and the occasional user of small amounts of census data. It is to these services that a user should go for data in hard copy or even diskette form. In addition, I should point out that six years is a very long time in terms of changes in computer technology, particularly in terms of changes in microcomputer technology. I have reviewed this recommendation with colleagues in both the computer field and the library field and they do not find it premature. I view CD-ROM primarily as a substitute for microfiche which will provide cheaper, easier to use copies of improved quality. In addition, proposals before the Joint Committee on Printing concerning providing depository libraries with data in electronic format will add further incentive to the library acquisition of CD-ROM readers and appropriate printers. The law covering depository libraries will enable users of these facilities to acquire census data at the cost of reproduction.

2. I realize that the testimony I provided on census data for racial and ethnic minority groups was ambiguous. I believe that the 1990 census should include the same number of questions about these groups, that the same amount should be published for governmental units and for larger statistical or other census-designated areas. My only recommendation was that we might, perhaps, increase the minimum racial or ethnic population in census tracts, e.g., for which we publish

tables. Since the average size of a census tract is about 4,000 people I thought we might up the minimum from 400 to 800, i.e. from 10% of the population to 20% of the population. This would save a nontrivial amount of paper and the data would be available for special tabulations. I do not believe the data for the smaller ethnic and racial populations is sufficiently used to warrent its being available in hard copy.

Again, I thank you for the invitation to testify. I trust that you have found my testimony of value in evaluating the Bureau of the Census' proposals for providing public access to the data from the 1990 census.

Sincerely,

Judith S. Rowe

Judith S. Rowe
Associate Director for
Research Services

JSR:BWJ

Enclosure

Mr. ACKERMAN. We will now hear from Dr. Benjamin F. King, the director of survey methods, Educational Testing Service.

Dr. King, welcome.

STATEMENT OF BENJAMIN F. KING, DIRECTOR, SURVEY METHODS, EDUCATIONAL TESTING SERVICE

Mr. KING. Thank you, Mr. Chairman.

Although seemingly a prosaic technical issue, the compilation of accurate lists of addresses for the structures that contain the housing units and other quarters where the population resides is the very backbone of census operations. A good list is essential for good coverage, and good coverage is in turn an important factor in achieving a successful enumeration.

The several tests of address list compilation activities discussed today are, I believe, excellent examples of the type of quasi-experiment that is most appropriate for the precensus years. Although there might possibly be a higher motivational effect when U.S. Postal Service employees are engaged in the real thing, I think that to be unlikely. Most importantly, there is no major element of cooperation by the general public that is required for full success of the experiment. I contrast this kind of situation with those of other tests of census procedures involving public response. An example is the recent test in Jersey City of the two-stage census procedure. As you probably know, the response rate in Jersey City was, I believe, 38 percent. There was a similarly low response in Tampa. That is, to the request for mail return, about 55 percent, I believe; and when this sort of phenomenon occurs, it is very difficult to interpret the results of these tests because the issue under consideration depends so much on public cooperation.

And, as an aside, I think that more of those kinds of tests should be embedded in the decennial census because without that the conditions are just not sufficiently realistic to provide easily interpreted results.

Back to the address list compilation procedures. The summary report is based on four preliminary research and evaluation memoranda. These are referred to as PREM's, and they are PREM No. 28, entitled "Results and Analysis of the Urban Address List Compilation Test," and PREM No. 38, dealing with the rural address list compilation test; PREM No. 12, unit by unit precanvass findings; and finally PREM No. 6, the "Results of the Advanced Post Office Check [APOC] II in the 1985 Pretest."

The order in which I just read these is the order in which I will briefly discuss them.

The impetus behind the 1984 compilation tests, both urban and rural, was the GAO report entitled "A $4 Billion Census in 1990? A Timely Decision." And that report recommended the investigation of the use of lists compiled by the U.S. Postal Service as the starting point for the sequence of update procedures that results in the final census frame. With respect to the basic design of the urban test, I find the Bureau to have been quite thorough in comparing the various combinations of starting lists—the vendor lists, the Post Office lists, the 1980 census list—and update methods—dependent canvass and the casing check.

I agree that it does not make sense to combine the casing check, which is a U.S. Postal Service operation, with the Postal Service starting list in laying out the experimental design. Furthermore, the Bureau should be commended for its frank discussion in PREM No. 28 of the shortcomings of the test, and I think the Bureau is very open about this.

For example, it clearly acknowledges the restriction of inferences to the two purposively selected test sites, Hartford and Bridgeport. This is always a problem in these tests. It would be wonderful if we could do the test in many different cities, but we have to select one or two. They are usually selected, or it seems at least, for worse-case characteristics, but unfortunately they turn out to have best-case characteristics for other aspects. So it leads to difficulty in interpretation.

With all of this openness and candor though, I am puzzled by one statement. This is in section I.F. of the PREM 28. I quote, ". . . this is not to imply . . . that there is not interest in knowing how these results would compare in other areas; areas with a higher growth rate, for example. On the contrary, the Bureau is presently examining ways in which this can be accomplished." That is the end of the quote.

If this statement is implying that there is some way to achieve this without actually testing the other areas, I should be very interested to find out how that is to be accomplished.

Anyway, in spite of the weaknesses, the results of this test are compelling and conclusive. Although the USPS list has high initial coverages compared to the vendor list, the cost after updating is so great that it cannot compete with the other two methods of listing. It appears that in its September 1986 decision, coming up, the Bureau will choose the basic approach used in 1980, that is, vendor lists followed by an Advance Post Office Check and then followed by dependent canvass, casing, and time of delivery checks. There are three Post Office checks and one census dependent canvass. The update procedures, however, are likely to be improved according to the findings from the 1985 and later census pretests.

In areas with poor vendor coverage, as Dr. Keane mentioned today, the address list from the 1980 census will serve as the most cost-effective start, and this is particularly relevant for certain intercity areas, for example, where the vendors themselves have no intention of going in and trying to create lists, at least not in the near future.

In my opinion these implied decisions are good decisions, but the Bureau also properly observes that the 1980 census list will not be as adequate in 1990 as it was in 1984 because of the obvious housing changes during the intervening 6 years.

Briefly, in the rural tests in Texas and in Georgia the Post Office listing was clearly inferior and cost ineffective because of errors in geography and map-related problems. The Bureau sees no chance of the USPS procedures being sufficiently improved nor of costs becoming sufficiently lower in time for the 1990 census and I concur with this judgment.

I also concur with the Bureau's optimistic view concerning the positive returns to be expected from continuing collaboration with the U.S. Postal Service as we move into the era of automated address files and automated navigation systems for enhancement of geocoding.

Now a brief word about the unit-by-unit precanvass. As described in PREM No. 12, the new precanvass procedures that were tested in the 1985 census in Jersey City and Tampa appear to have resulted in better coverage of individual apartments in multiunit buildings. In 1985 the enumerator did a unit-by-unit canvass within each apartment building whereas in 1980 information on number of units was usually obtained from a manager or some similar source. The additional cost of the unit-by-unit canvass would appear to be exceeded by the gains in correct apartment identification. Another finding is that the majority of apartment corrections in the unit-by-unit precanvass are for addresses that were classified in the Advance Post Office Check as deliverable without needing corrections, and this apparently confirms the need for a portfolio of update procedures rather than reliance on a single method.

There was another test—I mentioned the Advance Post Office Check II—in Jersey City and in Tampa, and with respect to experimental design this is the most interesting because of the clever method of salting addresses that were classified in the original APOC as undeliverable with a random sample of deliverable addresses to see if the U.S. Postal Service is really doing its job on checking on the original classification. It is encouraging to see that the results show no evidence of widespread Postal Service carelessness or rubberstamping in the second check.

The findings indicate, however, that the classification of an address as undeliverable in two checks is not sufficient evidence to delete the address, and thus the second Advance Post Office Check does not appear to be required. In other words, no matter what it accomplishes, it still is not enough to convince that an undeliverable is undeliverable.

Unfortunately, further analysis from 1985 of the combination of precanvass and Post Office determinations of status still does not indicate that addresses can be deleted before the followup checks. Less conservative reviewers might not agree with the implied Bureau decisions on these matters, but I think that caution in making dramatic changes to the present procedures is appropriate in this case.

In conclusion, I believe that in its 1986 and future activities the Bureau is pursuing a vigorous program of testing methods to improve address lists. The continued use of the unit-by-unit precanvass in Los Angeles County, the Advance Post Office Check and reconciliation in rural Mississippi, and the test involving an automated address list in one Mississippi County are high priority items. I believe that they are in accord with the recommendation of the Committee on National Statistics Panel on Decennial Census Methodology that scarce resources for testing need be applied to only the most promising coverage techniques. In other words, these are promising and they should be worked on.

I also look forward to the test in 1987 of the use of vendor lists as the starting point for address compiling in rural areas, which depend now on census prelisting. If my remarks today seem to be a blanket endorsement and excessively laudatory, it is only because Bureau personnel have been so thorough and innovative in their planning and execution of these particular tests. I might not feel the same way about some of the other tests, but I am confident that the decisions to be made concerning 1990 address list compilation procedures will be well founded and well reasoned.

Thank you.

Mr. ACKERMAN. Thank you very much, Dr. King, for your remarks and testimony.

[The response to written questions received from Mr. King follows:]

May 16, 1986

The Honorable Robert Garcia
United States House of Representatives
Washington, DC 20515

Dear Mr. Garcia:

This letter is in response to your letter of May 5, 1986 in which you thank me for testifying at the May 1 hearing on processing the 1990 Census. I, in turn, should like to thank you for the invitation and for the honor and privilege of appearing before your subcommittee. I hope that my testimony will be of help in your analysis and decision-making concerning Census operations.

With respect to the questions in the attachment to your letter, my response is as follows:

1. I examined all of the documents concerning the address list compilation tests that were sent to me before my testimony, and in addition, I read the summary report of those activities that was the basis for presentations by Bureau personnel at the recent Census Research Conference and at the April 1986 meeting of the Census Advisory Committee of the American Statistical Association. Assuming that the description of the results and the cost factors cited in those documents are accurate, I concur completely with the conclusions of the Bureau concerning the relative effectiveness of the various methods of compilation. I believe that the Bureau was fair and objective in its design and evaluation of the several experiments, and that full opportunity was provided in those experiments for the demonstration of the advantages of new methods of compilation if such advantages did in fact exist. To repeat a point that I made in my oral testimony, the address compilation tests were ideal for precensus execution because the results did not depend on public cooperation.

2. In contrast to the address compilation experiments, other precensus tests-- for example, the test of the two-stage census methodology-- depended heavily for their successful execution on the public's behaving much the same as it would have in a full-scale decennial census. In the 1985 Jersey City Test Census the mail return rate was only 38 percent, very much lower than the rates experienced in the worst districts of the 1980 Census. The low level of cooperation was probably due to a number of interacting factors, among them the fact that the operation was simply not "the real thing," and thus the spirit of participation in a national ceremony was absent. With the poor response rate in general, proponents of the two-stage approach to data collection will be able to criticize the Bureau's conclusion of failure of the method on the grounds that conditions were not sufficiently representative of true Census operations-- implying that the proposed two-stage method did not have a fair chance to display its advantages.

I do not agree with the above-mentioned critics of the Bureau's conclusions, but my reasons for disagreement lie in a strong prior belief that the two-stage process would not work, not in any great reliance on the results of the 1985 Census test. The Panel on Decennial Census Methodology of the Committee on National Statistics, of which I am a member, recommended in its 1984 interim report that the experiment with two-stage operations not be included in the 1985 Jersey City pretest, and that scarce resources be applied to other tests for which the outcomes were less certain a priori. The point that I wish to make here is not that the Census was remiss in its execution of the test, once the decision was made to include it in the 1985 operation, but rather that it was not a good idea to do the test in the first place because of the high prior probability of failure; and if one does not agree with that viewpoint, that the test should have been embedded in the Decennial Census rather than in a precensus operation with low likelihood of widespread public cooperation. It follows that this last point of criticism applies to any test for which unambiguous interpretation of the results depends on a high level of public compliance with the request for data.

3. With respect to the last question about possible Census use of other government files to reduce its task, I have done very little research on this matter, and I can only give you an opinion based on my general experience in studying Census operations and on my professional experience as a teacher and practitioner of statistics. The essence of statistical estimation and inference in many practical applications involves the use of information on certain variables to reduce uncertainty about the values of other variables of interest. The Census task of counting and estimating the characteristics of the population is no different in that regard, and thus, if there were no concerns about confidentiality, the use of information from other governmental agencies would be of great value in attaining a better count, as well as more accurate measures of person, household, and family characteristics. Family income, for example, would certainly be better measured if the files of the Internal Revenue Service could be used routinely to provide that information instead of having to elicit it in the long-form Census schedule; and there would be considerable ultimate cost savings, as well as reduction of burden on respondents.

I would be the last person, however, to suggest that such a merger of data could be easily accomplished without raising public suspicion about invasion of privacy. Yet, if the public believes that the information that it provides to the Bureau is adequately protected from disclosure at the personal level, and if it places similar trust in the IRS, there is no reason in principle that it should not place equal trust in a joint Census-IRS use of the same data. The problem is one of figuring out how to achieve that level of trust, while accomplishing the desired end-- i.e., more accurate measurement. I do not have the solution. I earnestly hope, however, that the attitude in the future of all branches of government--executive, legislative, and judicial-- will be to try to find ways to facilitate the merger of agency files for the purpose of more efficient and accurate measurement, while protecting our rights to privacy, rather than taking the position that any cross-agency use of files will necessarily lead to damage of our rights.

Sincerely yours,

Benjamin F. King

Benjamin F. King

Mr. **Ackerman**. As the Chair previously suggested, I would like to ask Dr. Keane to join the table at this time and to share with us some of his thoughts and responses to the testimony that we heard, and the Chair would like to advise that we do have a limited amount of time, so we will keep the free-for-all down to a minimum.

Mr. **Keane**. Thank you for that opportunity, Mr. Chairman.

I should identify my colleague, Peter Bounpane, who is known to this subcommittee because he usually accompanies me to hearings like this, and I am glad he is here. I would like him next to me, but we cannot have everything.

Mr. **Ackerman**. We could arrange that. Peter, if you would like to move your chair to the other side.

Mr. **Keane**. I will start talking while they are moving, and both of us would like to comment so that may help the future discussion, and we will be to the point. Mine are a bit more generic than Peter's would be.

I would like to talk, first of all, about the size and scope of an activity the size of a decennial census in this country, and we have quite a bit of comparative experience over time as well as with other countries. For instance, in Saudi Arabia there are about 10 million people. I just came back from Egypt where they are doing a census of about 50 million people. When you are talking about a quarter of 1 billion people—quarter of 1 billion people—it is a new contest. Things that work in large surveys do not necessarily work here.

For instance, I had a conversation with American Telephone & Telegraph—predivestiture. They mail—and this is relevant to us—3.1 million annual reports. When I told them about the Census Bureau mailing 30 times that many, how would they handle it, with what kind of technology, their eyes rolled. So as high technology an organization as AT&T would not know how to handle the kind of mailings that faced us in the past and will face us in 1990.

To elaborate a little further, when we are talking about something like a census done for a tenth of the country over a 10-year span, we immediately get into cost considerations that are already sensitive now, that could be extraordinarily so with the kind of money and budgeting involved; and there is also an issue of the constitutionality of doing something like that. And we know how difficult it is to make an amendment to the Constitution. So those are the kind of considerations.

We are very much enriched by this kind of a hearing and these kinds of comments. All of the individuals here are known to us and we value their comments.

I would tack on four considerations when it comes to automation of the 1990 census. One is the hardware and particularly the installation and maintenance obstacles, challenges posed in the decentralized kind of mode which we discussed. Related to that is software development, enormous software development, the kind of leadtime that requires; and also the acceptable space, trying to acquire that in the places and in sufficient amounts that are affordable to us. These things have not been mentioned, the space, for instance. The staffing has and that is the final of the four points.

To get the number of managerial and technical people in the locations needed, to get them trained, to get them in synchronization with each other and with the other four areas pose monumental challenges.

So I summarize by saying this is the largest Federal Government program in the sense it touches more people than anything else this Government does, and that means that it is in a special category by itself. Things that work elsewhere in other large surveys or other countries or in past times do not necessarily work here. That is why if we seem a bit cautious at times it is with good reason.

If I may turn to my colleague. Thank you.

Mr. ACKERMAN. Thank you, Doctor.

Mr. BOUNPANE. Thank you, Mr. Chairman.

I am not going to address all the issues I heard raised, in the interests of time, but let me just pick a few I thought were key—and some that recurred among the four—and try to say just a few words about those.

The first is the general question of "Have we gone far enough in automating 1990?" Picking up a little bit on what Jack just said, this is a very difficult question and, obviously, people have different opinions about that particular issue. To some, it may seem our eventual choice was perhaps too conservative, that we could have been further along the continuum than we actually are.

In some respects the answer to that is, "Yes, we were physically able to be a little further along that continuum." There are machines that can do the kinds of things that people have said here this morning. The question is can they be purchased, implemented, tested and work properly in a huge one-time activity like the census? It comes very quickly. It is over very quickly, and then it is done. You have to do something with the equipment at the end of the census, and our eventual choice considered those factors as well as what is physically available on the market at the present time. And there are differences of opinion about this. We think we took the choice that maximizes the use of equipment available to us but gives us the highest probability of no failure in 1990. That would be very bad, to have a fully automated census, be relying on it, and push the button in April 1990 and have it not work.

We tried to balance that risk against the risk of not going far enough. This difference of opinion that exists here at this table also exists with many people within the Census Bureau. It is a difficult choice.

One thing that should also be mentioned here is our reluctance to use multiple systems within the census, and in general that is correct. We like to have one way to do things and to use that one way throughout the country. That is because we have to hire a large temporary work force, train them quickly and ask them to do this job. The more exceptions you have to the rules, the more difficult it is to manage. And so as Bill Eddy pointed out, it would be very possible to collect certain information over the phone, have the respondent call in rather than fill the questionnaire in by pencil. We were convinced that not enough people could do that to make that the sole collection technique and, therefore, we use paper. We do, however, allow respondents to call in in some instances, perhaps not to the extent he thinks we ought to.

A question that was raised by Dan Horvitz was, "Should we use more computer-assisted telephone interviewing?" Here again, once we had to develop a system to use the paper questionnaires, the overhead costs of introducing a whole new CATI system just to handle those people who failed edit—that is, did not answer all the questions they were supposed to—was pretty large. We judged that it is not worth that investment for that small a universe, not that we do not want to use telephone interviewing and not that the benefits that were pointed out are not really there.

Our question is, "Can we get a technique in place and manage it in a large-scale census?" We are not always so stubborn on having only one system. Let me point out one example already raised here and that is, "How do you check in the questionnaire with a bar code?" As we pointed out, it could be done one of two ways: A laser sorter, which is simply a machine that reads bar codes, or a simple wand run across the bar code. Laser sorters are extremely fast. As many as 30,000 questionnaires an hour can go through one of these machines. It is very helpful, but they are very expensive. No afterlife after the census is a major consideration. Using a small personal computer with a wand attached is much slower, but that kind of a machine has tremendous use after the census, and so we struck a balance. We have decided to use laser sorters in a chunk of the country, where we think we need to check in very fast, but, in the balance of the country we will use small computers with a wand, and they will have tremendous use after the census. So occasionally we do bend this rule.

Also I should point out——

Mr. ACKERMAN. Could I ask a question at this point concerning the economy. I think I hear from what you are saying that the preferable method would be the laser use except for the fact that there is a one-time cost and no afterlife. Would it not be possible for these to be leased or for arrangements to be made to buy and resell them afterward? Has anybody investigated the cost comparison that way?

Mr. BOUNPANE. Yes; we have looked at that. So far we have not been able to find them to be leased. Perhaps it is possible, but no one has been willing to lease them to us because they have to be built to order; and that being the case, we have not found an after-census purchaser for them at this point in time.

I might point out, Mr. Chairman, laser sorters cost about a quarter of a million a machine.

Mr. ACKERMAN. How many would you need?

Mr. BOUNPANE. Depending on how you did this, as many as perhaps 30, 40.

Mr. ACKERMAN. That is $7½ million, if my math is correct.

Mr. BOUNPANE. Yes.

Mr. ACKERMAN. What does the labor cost of all those people with the magic wands come to?

Mr. BOUNPANE. I do not know the answer. It is probably more.

Mr. HORVITZ. According to the number the Bureau presented in the technical reports associated with this area, the cost trade-off was much in favor of the wand and personnel rather than the laser sorters, so I do not understand the Bureau's still strong interest in laser sorters.

Mr. **Bounpane**. The reason we are interested in laser sorters is that, in certain parts of the country, it is very difficult to get enough people to do this job where we have to do it in a short amount of time, and it is better to use the available people for other things. For example, in an area like New York, where we want to hire enough people to make sure we have enumerated everyone, it might be very beneficial to have a laser sorter there so we can free up as many human resources as possible for the most difficult task, which is the enumeration. That is the kind of reasoning we went through. In some parts of the country it makes sense.

Mr. **Ackerman**. Anybody is free to respond if you want at any time.

Mr. **Bounpane**. I thought the suggestions on the year 2000 were very good, and we will certainly look into those. It seems like a long time between censuses and there should be plenty of time to plan for the next one, but our experience is generally there is not. Even though we do have test censuses, it is hard to learn everything we would like to learn to plan the next one, and so we should probably make more use of each census to learn for future censuses.

Many of the suggestions that were made we would very much like to try in the 1990 census to see if it is possible to use them in future activities. So we are going to consider nonpaper collection technologies, allowing people to call in and use the computer, and optical character recognition. Tthere seems to be some problem with OCR when you have handwritten letters as opposed to just numbers, but at least we will experiment with that in 1990 so we will have information on those technologies to use in planning for the 2000 census.

The issue of transmitting data over the phone lines came up in several people's remarks, and we have looked at that. That is definitely possible, to transmit information over phone lines from decentralized locations to centralized locations. We have decided, however, not to do that for two reasons: The first reason is our examination of the process says that moving the data is not our real hangup—at the point we actually have the data collected, moving it physically is not our real problem. The second reason, which is also very, very important, is that we were very worried about putting data into phone lines, even with encryption devices, since that might allow someone to intercept it. It would only take one person to do that to cause a significant problem. The alternative is to hard-wire with our own dedicated lines, but the cost of that would be just enormous.

A couple of other thoughts. I thought Judith made some very good comments about the products, and I would say that our planning is definitely in the direction of more CD-ROM type activities and less micro, floppies and microfiche. Although we get stuck in a dilemma here, that our products must serve a wide range of users and those of us in the room are very familiar with electronic devices, but there are people in the United States who are not at the same level and they want census data too. So again we have to find a compromise which may not be as high a level as some may wish.

And a couple of words on the edit, because that also came up fairly strongly, that it is disappointing to some that we are not doing an automated edit throughout the entire country. We would also like to have done that. We agree that a hand edit has difficulties. That is why we moved towards concurrent processing. We would like to have been able to do automated edit nationwide. Our judgment is, however, we cannot accomplish that.

The point I would like to make is that even in those areas of the country where we are planning a hand edit we also plan to do a computerized, automated edit later at a more relaxed time period.

Before I close, I would like to elaborate on something Ben King said. I know it is not the purpose of this hearing to talk about public cooperation, but Ben made that point and I think it is a very good one. These are good issues here and we should talk about them and should resolve them, but the census is eventually going to be dependent on people's willingness to cooperate with it, and that is an important issue and we would like to spend as much time on that as on some of these others.

These are some of the thoughts I jotted down as I listened to the testimony. Thank you, Mr. Chairman.

Mr. **Ackerman**. Thank you very much.

Dr. Eddy.

Mr. **Eddy**. I would like to speak to the optical character recognition issue just briefly. Several years ago I was a consultant to a USPS vendor and they were at that time delivering to the USPS sorting machines that used optical character recognition technology, and it is my understanding that they were able to not only read the ZIP Codes but actually city names and states and they were able to look up these in various tables and so forth at rates comparable to the rates for laser sorters that Peter just mentioned.

I think it is probably impossible to implement this sort of thing for 1990, and I would not encourage it; but by the year 2000 I would presume—I do not want to say it is trivial, but it certainly should be quite doable by then, and I would like to strongly encourage it.

On a different note, I am really a little concerned with the rapidity with which the Census Bureau wants to dismiss the notion of electronic transmission of the data. They may or may not be aware that there are other organizations in this country which are more concerned about security than they are, and they seem to be able to make use of the public telephone network. I am thinking not only of various federal government agencies such as the Department of Defense, but rather public corporations such as commercial banks which transmit hundreds of millions of dollars through the public telephone system on a regular basis, and they do use encryption. There are various encryption standards; and if they are good enough for banks and for the Department of Defense, I am certain they are good enough for the Census.

Thank you.

Mr. **Ackerman**. Thank you, Doctor.

Mr. **Bounpane**. Just to add one thing, that is a reasonable point and I did not want to leave the impression we were doing nothing on phone lines. We do transmit certain information and are doing some experiments on some others. The question was whether the whole data file should be transmitted over phone lines.

Mr. **Keane**. I have a comment of Professor Eddy's last point. If you are referring to the intelligence community as someone else, for instance, they probably do not have our mandate which is to provide their data to virtually everybody. It might be quite the opposite. So that is one consideration.

Another consideration is those commercial firms that do use the electronic dissemination and how far we go as a supplier and how far they would like to see us go and what the relative efficiencies are, that that is something that from a strategic planning standpoint the Census Bureau is concerned with right now.

And, finally, there is always a cost consideration in the sense of what is the most efficient medium and, using all the media with which we might disseminate data, how do we get a complementary balance.

Mr. **Ackerman**. Yes, Dr. Horvitz.

Mr. **Horvitz**. I just wanted to comment that I had made a suggestion that it is time to consider some census alternatives than one that mobilizes every 10 years. It seems to me the comments we heard here from Mr. Keane and Peter Bounpane suggest, in fact, the problem of logistics is a very real problem in organizing the census and determining where the offices are to be located, how many offices, what kinds of people to hire, the problems of training those people. Those are major issues, and they seem to stand in the way, in fact, of certain decisions that would move the Bureau closer towards a level of automation that I certainly feel was achievable for the 1990 census and which from my standpoint now is not going to be even approached.

Mr. **Ackerman**. Thank you very much.

Mr. **Horvitz**. I have not finished. I appreciate the opportunity to just say a few more words.

Mr. **Ackerman**. In that case I won't thank you for a while. [Laughter.]

Mr. **Horvitz**. Jack Keane raised the issue of the apportionment requirement, and I very readily recognize that that is an issue. What I am suggesting is it is time to look at alternatives to what we are doing and have been doing since 1790. I do not think that tradition ought to necessarily hold in a world that is changing with technology and in other ways so rapidly as it is.

Now, if there is a census alternative and it stands up to all of the requirements that one would put on a census, then it seems to me that given that census alternative it is time then to examine the issue of apportionment and whether, in fact, there is an alternative to our current ways of deciding apportionment. I certainly would not suggest we abandon what we are doing now in terms of apportionment, but in the face of a decent alternative I think we should seriously examine that issue.

So I think really that the Bureau ought to find what are the best alternatives for the United States and in the public interest and leave the problem of apportionment to the Congress, given the pressure that they would be under in the Congress to consider the alternative.

I think that is all.

Mr. **Ackerman**. Thank you.

Let me thank the Director and the entire panel for their participation with us this morning, for the expert testimony, for the responses, for the presentations on behalf of the committee and especially Chairman Garcia. We would like the participants, if they would be willing, to be prepared to answer in the future any questions that the chairman or the subcommittee might direct to you so that we may have those included in the record as well. Thank you all very much.

To summarize the experts' remarks and to react to the Bureau's entire plan, we have Prof. Stephen Fienberg who is the Maurice Falk Professor of Statistics and Social Science at Carnegie Mellon University. Dr. Fienberg is the current chairman of the Committee on National Statistics of the National Academy of Sciences, and as such he has been instrumental in reviewing the work of the Census Bureau conducted by the Academy. He is also vice president of the American Statistical Association.

Dr. Fienberg, welcome.

STATEMENT OF STEPHEN E. FIENBERG, MAURICE FALK PROFESSOR OF STATISTICS AND SOCIAL SCIENCE AT CARNEGIE MELLON UNIVERSITY

Mr. **Fienberg**. Thank you, Mr. Chairman. I was getting lonely back there being the only witness excluded from the table.

It is a pleasure for me to appear before the subcommittee and to participate in its review of the Census Bureau's plans for 1990. I have been following their planning effort now for some time with great interest, in part because of my personal research activities and in part because of my activities as chairman of the Committee on National Statistics. The committee has had a special independent panel commissioned by the Bureau which has been examining the methodology for 1990. The panel issues a report last fall entitled "The Bicentennial Census: New Directions for Methodology in 1990," and we made copies available to the subcommittee staff and to the subcommittee members. Some of the panel's comments and recommendations are germane to the topics that care the focus of today's hearing as well as to the subcommittee's ongoing responsibilities.

My comments today are going to be focused in three areas: One is just a reminder about the extent of the long-range aspects of the census planning; the second is going to be an attempt at a very broad and sweeping summary of the remarks and testimony of the other witnesses today; and then finally I want to mention additional areas of census planning for 1990 that I believe require outside scrutiny and comment as well as congressional oversight and guidance.

Information gathered as part of the decennial census is used not only for purposes of congressional reapportionment and for State and local redistricting but also for the distribution of billions of dollars of Federal funds and for a host of other government and nongovernment purposes. With so much riding on the outcome of the decennial census, we should not be surprised at the extent of planning required. Census taking is a massive enterprise. Planning for 1990 officially began in the fall of 1983 with the appropriation for fiscal year 1984, but the real planning began much earlier, well before the 1980 census was actually taken. That planning program included several experiments and post-enumeration studies designed to help develop and improve methodology for subsequent censuses, including the 1990 census.

The need for such long-range planning efforts underscores both the strength and the weakness of the Bureau's current activities.

Mr. **Ackerman**. Dr. Fienberg, let me just, if I might, advise you there is a vote presently in progress so that you can properly pace yourself and we would not have to adjourn and reconvene. You are going to have a total of about 7 minutes.

Mr. **Fienberg**. I will be done.

Today's hearing focused on four very specific technical aspects of the methodology for 1990, and in each area you have heard about the extent of the long-range plans of the Census Bureau staff and the step-by-step evaluations in which they are currently engaged. These efforts are consistent with the Census Bureau's publicly stated minimum goals which are: first, to conduct the census without increasing the per housing unit cost in 1980 dollars; second, to expedite the availability of the data to users; and, third, and I want to emphasize the third one, to maintain a high rate of overall coverage and to improve the accuracy of small area data.

Improved address lists and automation are two of the keys to the first two goals I enumerated, but by focusing on them the Bureau has implicitly assumed these will help achieve goal three.

Today's presentations have not really provided any direct support for such an assumption. I continue to be concerned about overall coverage, undercoverage in selected population groups and areas, and the accuracy of small area data.

The outside experts who examined the Census Bureau plans on processing procedures and commented on them this morning have all praised the Bureau for its efforts, and I concur in that praise. But some of these experts have also expressed concern that the 1990 census will be done with 1970 and 1980 technology. That is, there is too much satisfaction with the status quo and not a rapid enough movement toward techniques and approaches already in widespread use outside the Bureau.

As an example, I would take the IBM PC/XT, the personal computer that is a cornerstone in some of the automation procedures. We acquired several of those a few years back in my department at Carnegie Mellon and they have since been discarded for new computers. The IBM PC/XT that was sitting in my office is used primarily by my children rather than by the professionals working at Carnegie Mellon University.

These experts have also expressed the need for research on new approaches to census taking if we are to have quality and cost effective population data in the future. I note that comments have much in common with the recommendations of the committee's panel. In many ways the panel took as its starting point the structure for the 1990 decennial would resemble that for the 1980 and the panel supported the Bureau's approach to address list compiling and the variety of automation procedures that we heard described today.

The panel then focused on four other topics: Plans for research and ongoing experimentation with new methodology in 1990 and beyond; improvements in questionnaire content and evaluation of questionnaire changes; coverage improvement evaluation; and estimation and the adjustment of census data for undercount and overcount. Each of these areas is critical to the successful completion of the 1990 census, and each has been the focus of research and evaluation activities at the Bureau.

If you consider, for example, the problem of estimation and adjustment, you will see the nature of the concern and the importance. The Bureau itself has raised the issue of whether the census ought to be viewed as a count, an estimation effort, or some combination of the two. Many of us outside the Bureau have come to think about census taking as a statistical estimation problem, but we agree with Census Bureau staff that the census effort should be some combination of estimation and counting.

The issue is what combination? Even though census data will not be gathered for another 4 years, the Bureau needs to set in place in the near future, that is, in the next year or two, operational versions of adjustment procedures if the full value of census adjustments and estimations is to be realized. This means a firm timetable must be established that allows for outside scrutiny and comment and for congressional oversight and guidance.

Prof. William Kruskal, a former colleague of mine at the University of Chicago, remarked that: "The decennial census is a national ceremony and a symbol of the relationship between citizen and government." But if we are to go beyond the ceremony and the symbolism, we need to look carefully at the uses of decennial census data and we need to make a professional evaluation of the methodology for data acquisition and analysis.

This subcommittee's ongoing review and oversight of the Census Bureau plans are critical in the decisions on methodology for the 1990 census. These decisions will help to shape the information base that will guide Government policy into the 21st century. It has been my privilege to help the subcommittee in this effort.

[The statement of Stephen E. Fienberg and his response to written questions follow:]

May 1, 1986

STATEMENT
OF
STEPHEN E. FIENBERG
MAURICE FALK PROFESSOR OF STATISTICS AND SOCIAL SCIENCE AT CARNEGIE MELLON UNIVERSITY
TO THE
SUBCOMMITTEE ON CENSUS AND POPULATION
OF THE
COMMITTEE ON POST OFFICE AND CIVIL SERVICE
U.S. HOUSE OF REPRESENTATIVES

Mr. Chairman, it is a pleasure for me to appear before your Subcommittee again, and to participate in your review of the Census Bureau's plans for the 1990 Decennial Census. I have been following this planning effort with great interest, in part because of my personal research activities, and in part through my position as Chairman of the Committee on National Statistics, at the National Academy of Sciences. The Committee has had a special panel, commissioned by the Bureau of the Census, which has been examining the methodological planning for 1990. The Panel issued a report last fall entitled: *The Bicentennial Census: New Directions for Methodology in 1990*, and some of the Panel's comments and recommendations are germane to the topics which are the focus of today's hearing, as well as to topics that may be the subject of future hearings.

My comments are divided into three parts: the first part is a reminder of the extent and long-range aspects of census planning; the second part is my attempted synthesis of the remarks and formal testimony of the other witnesses; and the third part focusses on additional areas of census planning for 1990 that I believe require outside scrutiny and comment *as well as* congressional oversight and guidance.

CENSUS PLANNING

The information gathered as part of the decennial census is used, not only for purposes of congressional reapportionment and for state and local redistricting, but also for the distribution of billions of dollars of Federal funds, for a host of other Federal government needs, and for a variety of purposes by researchers, planners and decision makers in business, state and local governments, and academic institutions. With so much riding on the outcome of the decennial census, we should not be surprised at the extent of planning required -- after all, census taking *is* a massive enterprise.

Planning for the 1990 Census officially began in the fall of 1983 with the appropriation for fiscal year 1984, but the real planning began much earlier, well before the 1980 census was actually taken. The planning program for that census included several experiments and post-enumeration studies, designed to help develop improved methodology for subsequent censuses including that of 1990. The need for such long-range planning efforts underscores *both* the strength *and* the weakness of Bureau of the Census's current activities. As was noted by the Panel on Decennial Census Methodology:

> To the general public and many casual users of census data, it may appear that the Census Bureau has ample time to plan wisely for the 1990 Census, given the start of the planning process more than six years prior to Census Day, April 1, 1990, and the foundation of research already completed in connection with prior censuses. In fact, as a review of the Census Bureau's field test schedule for 1990 indicates, there are relatively few opportunities to test thoroughly changes or modifications to census procedures, particularly if the changes represent major departures from the past. Moreover, evaluation of the likely impact of important changes is hampered by the fact that pretests cannot adequately assess the effects of alternative procedures on public cooperation with the census -- only tests conducted under census conditions, that is, experiments incorporated into an actual census as distinct from pretests, can fully address this important question.
>
> . . .
>
> In addition to the compressed time schedule for testing and research, two other critical factors affect the ability of the Census Bureau to modify census methodology: staff and budget resources. The Census Bureau has long been known for the high quality and dedication of its technical staff. The current budget for research on decennial census methodology, particularly for research on the undercount, is large by the standards of earlier censuses. Nevertheless, no agency of government, particularly in the constrained world of the 1980s, can expect to have sufficient staff or resources to try out more than a few promising ideas and concepts. Pressures in the next few years to reduce the federal government's large deficit may make it more than usually difficult to obtain adequate staff and funding to carry out a thorough research and testing program for 1990.
>
> *The Bicentennial Census* (1985, pp.4-5)

TODAY'S HEARING

Today's hearing has focussed on four technical aspects of the methodology for 1990: (i) address list compilation, (ii) automated address control file and automated check-in, (iii) concurrent data processing, and (iv) data products and processing. In each area, you have heard about the extensive and long-range plans developed by Census Bureau staff, and the step-by-step evaluations in which they are currently engaged. These efforts are consistent with the Census Bureau's publicly-stated minimum goals to

> (a) Conduct the 1990 Census without increasing the per-housing-unit cost in 1980 dollars. (b) Expedite the availability of the data to users. (c) Maintain a high rate of overall coverage and improve the accuracy of small area data while reducing the overall differential for population groups and geographic areas.
>
> Bailar (1984)

Improved address lists and automation are two of the keys to goals (a) and (b), but by focussing on them the Census Bureau has implicitly assumed that these will help achieve goal (c). Today's presentations have not provided any direct support for such an assumption, and I continue to be concerned about overall coverage, undercoverage of selected population groups and areas, and the accuracy of small area data.

The outside experts who have examined the Census Bureau's plans on processing procedures and commented upon them this morning have all praised the Bureau for its efforts. I concur in this praise. But some of these experts have also expressed concern that the 1990 Census will be done with 1970s and 1980 technology, i.e. that there is too much satisfaction with the status quo and not a rapid enough movement towards techniques and approaches already in widespread use outside of the Census Bureau. They have also stressed the need for research on new approaches to Census-taking if we are to have quality and cost-effective population data in the future. What follows is a brief summary of the comments of each of the four experts.

Testimony of Dr. Benjamin F. King

Dr. King's evaluation of the Census Bureau's compilation of accurate lists of addresses was quite laudatory. Such lists form the backbone of Census operations, and Dr. King found the results of recent tests supporting the use of outside vendor lists to be compelling and conclusive.

Testimony of Dr. Daniel G. Horvitz

Dr. Horvitz examined the Census Bureau's new approaches to automated address control files and to automated check-in. While he was supportive of the direction of current activities in this area, and the likely improvements in productivity and quality that they should produce, he expressed concern that Census Bureau staff were too tentative in their adoption of new computer-based technology. In particular, Dr. Horvitz noted the uses he would make in automated data collection and check-in procedures of Computer Assisted Telephone Interview (CATI) methodology. Dr. Horvitz also suggested a radically different approach to doing the census that would spread data collection over the decade instead of implementing a massive mobilization once every 10 years.

Testimony of Professor William F. Eddy

Professor Eddy also reviewed the materials on automated address files and check-in, and his remarks were complementary to those of Dr. Horvitz, encouraging the Bureau to make more effective use of computer technology. Professor Eddy strongly supported the Bureau's move towards concurrent processing in 1990, but also suggested that the Bureau move more boldly away from its plan to sort the paper forms. Professor Eddy concluded his testimony with some speculation of the impact of computers on census-taking in 2000 and beyond. Actually, part of this speculation is already a technological reality and should be influencing thinking about the 1990 Census.

Testimony of Mrs. Judith S. Rowe

Mrs. Rowe provided the Subcommitee with a detailed evaluation of the Bureau's plans for data products and dissemination. She supported the proposed schedule for release of data products, but expressed reservations about the Bureau's current thoughts on media for dissemination. In particular, Dr. Rowe suggested a continued movement away from massive printer reprints and especially away from microfiche. She agreed with the focus on the emerging CD-ROM or laser-disk technology and argued that laser disks should replace *both* microfiche *and* floppy disk technology. She supported the role of CENDATA as a dissemination approach for some classes of users. Finally, Dr. Rowe encouraged a continued focus by the Bureau on *timely* release of *accurate* data, and discouraged the Bureau from attempting to compete with outside groups and firms in the area of software development.

ON WHAT OTHER TOPICS SHOULD ATTENTION BE FOCUSSED?

I note that these comments from the outside experts have much in common with the recommendations and comments of the Committee on National Statistics' Panel on Decennial Census Methodology. In many ways, the Panel took as its starting point that the structure of the 1990 Census would resemble that for 1980, and the Panel supported the Bureau's approach to address list compilation and "its efforts to develop improved automated procedures that have the potential to speed up data collection, improve accuracy, and reduce costs." The Panel's report then focussed on four other topics:

(i) Plans for research on and experimentation with new methodology in the pretests and in the 1990 Census,

(ii) Improvements in questionnaire content and the evaluation of questionnaire changes,

(iii) Coverage improvement evaluation,

(iv) Estimation and the adjustment of census data for undercount and overcount.

Each of these areas is critical to the success of the 1990 Decennial Census and each has been the focus of research and evaluation activities both within and without the Census Bureau.

Consider, for example, the issue of estimation and adjustment. The Bureau itself has raised the issue of whether the census ought to be viewed as a counting effort, an estimation effort, or some combination of both. Many of us outside the Bureau have come to think about census-taking as a statistical estimation problem, but we agree with Census Bureau staff that the census effort should be some combination of estimation and counting. The issue is then what combination. Even though the Census data will not be gathered for another four years, the Bureau needs to set in place, in the near future, operational versions of adjustment procedures if the full value of census adjustment and estimation is to be realized. This means that a firm timetable must be established that allows for outside scrutiny and comment *and* for congressional oversight and guidance.

Professor Wm. Kruskal of the University of Chicago has remarked:

> The decennial census is a national ceremony and a symbol of the relationship between citizen and government. Whatever one's view of the census, whatever one's philosophical position about the Federal Government, it may be argued that the census is one of our relatively few national, secular ceremonies. It provides a sense of social cohesion, and a kind of nonreligious communion: we enter the census apparatus as individual identities with a handful of characteristics; then later we receive from the census a group snapshot of ourselves at the ceremony date. Like many family pictures, the snapshot is a little blurry in spots, but recognizable and fascinating to compare across the decades.
>
> These symbolic matters are not just poetic speculation. I believe that they play important roles in the actual carrying out of the Census, in congressional discussions of the Census, in beliefs that take extreme positions about the accuracy of the Census, and no doubt in other ways.
>
> Kruskal (1984, pp. 49-50)

But if we are to go beyond the ceremony and the symbolism, we need to look carefully at the uses of the decennial census and we need to make a professional evaluation of the methodology for data acquisition and analysis. This Subcommittee's ongoing review and oversight of the Census Bureau's plans is a critical component in the decisions on methodology for the 1990 Census. These decisions will help to shape the information base that will guide government policy into the 21st century. It is my privilege to assist you in this review.

RESPONSES BY STEPHEN E. FIENBERG TO QUESTIONS POSED BY CHAIRMAN GARCIA FOR THE RECORD OF THE HEARING OF MAY 1, 1986

1. **Question:** Please inform us where your plans stand in reviewing the Census Bureau's decennial plans. When will the panel meet next to review the Census Bureau's plans?

Answer: As you are aware the Committee on National Statistics' Panel on Decennial Census Methodology completed the first phase of its work with the report issued in September 1985. There had been tentative plans for a continuation of the Panel's activities, including intensive reviews of methodological planning and the evaluation of results from the 1985 and 1986 test censuses. The Bureau of the Census did not fund this second phase and only recently extended the contract for support of the Panel in a substantially reduced amount. This additional funding will cover a single meeting of the Panel, and we have tentatively scheduled this meeting for September, a full year after the Panel last met. We have also had preliminary discussions with Bureau of the Census staff regarding funding for the Panel for FY1987.

2. **Question:** Considering all of the work of the Census Bureau and its plans for the 1990 Census, what do you think are the two or three most important actions that this subcommittee could take to improve the census and help it to be more accurate, timely, and useful?

Answer. The two areas in which the Subcommittee can most effectively assist the Bureau of the Census are in the procurement of computer equipment, and in oversight of plans for adjustment of censal counts.

First I draw your attention to a recommendation by Dr. Horvitz in his written statement:

> "This Subcommittee, and the Congress, could assist the decision process at the Bureau by, first, demanding careful documentation of the computer hardware requirements to conduct an automated census at the planned level; second, based on this documentation, negotiating a mutually acceptable set of requirements; and third, providing some assurance to the Bureau that the separate budget line item for the computer hardware included in the mutually acceptable set of requirements will remain intact. I recommend completion of this process no later than the end of this fiscal year."

The Bureau needs to know what hardware it will use and it needs to develop the software for that hardware in order that it can be tested in the dress rehearsal scheduled for 1988.

Second, I would urge the Subcommittee to schedule a hearing to review plans for adjusting the census, and to set a firm schedule for decisions regarding adjustment. Adequate time must be allowed for professional and congressional input on the issue of adjustment. Attention to this issue by the Subcommittee can help the Bureau reach a timely and reasoned decision.

3. **Question**: Do you think that the Bureau's processing plans will make adequate allowance for evaluating the census coverage on time to report it when the apportionment figures are completed?

Answer: The Bureau's move towards concurrent processing should allow for census coverage evaluation in a more timely fashion in 1990 than was the case in 1980. Nonetheless, if 1980 is to serve as a guide, the Subcommittee is wise to be concerned about whether there is adequate allowance of coverage evaluation prior to the deadline for reporting of apportionment figures. After all, in 1980 the Bureau's demographic analysis initially showed a net overcoverage of the white population, a result which subsequent analyses showed to be incorrect. New approaches to coverage evaluation, if adopted, would offer greater assurance that an adequate evaluation would occur by the time apportionment figures are due.

The Panel on Decennial Census Methodology made several suggestions for improvements in the area of coverage evaluation, including the use of sampling for follow-up of nonrespondents and coverage and the use of systematic observation. The Bureau had rejected most of these suggestions for 1990. One final method for improving the timeliness of coverage evaluation is through the use of a pre-enumeration survey in place of the post-enumeration survey used in 1980. This approach is still under consideration by the Bureau.

4. Question: Please comment on the Census Bureau's internal report that suggests they will decide whether to adjust the census after all the figures are in.

- In your experience as a statistician, are there adequate procedures and agreed upon standards available to allow the Census Bureau director to make an unambiguous decision about adjustment after the results are known?

Answer: I have been told about the report to which you refer and I think that any plan that will defer the decision on whether to adjust the census until all the figures are available is sheer folly. No other action I can think of is more likely to invite a repeat of acrimony and law suits that followed the Bureau's decision not to adjust in 1980. This plan could destroy the credibility of the Bureau of the Census in other technical areas.

The basic methodology for adjusting the census using data from pre-enumeration or post-enumeration surveys is well-developed and widely-accepted. It is my belief that adequate procedures and agreed-upon standards are available or could be developed in the near future in order to allow for an unambiguous decision about adjustment in advance. I would urge that closure on the details of adjustment methodology should be reached in the near future and that a full-scale test be conducted in conjunction with the census dress rehearsal. A final and reasoned decision could then be made well in advance of the 1990 Census.

Mr. Ackerman. Dr. Fienberg, thank you very much for your remarks and your testimony. The subcommittee greatly appreciates it and will request that you as well, if you will, be prepared to respond to any questions that Chairman Garcia may direct to you for inclusion in the record. Thank you very much and thank you all very much for being here.

The subcommittee stands adjourned.

[Whereupon, at 11:55 a.m., the subcommittee was adjourned.]

CENSUS QUESTIONNAIRE AND AUTOMATION

THURSDAY, MAY 15, 1986

House of Representatives,
Subcommittee on Census and Population,
Committee on Post Office and Civil Service,
Washington, DC.

The subcommittee met, pursuant to notice, at 10:15 a.m., in room 311, Cannon House Office Building, Hon. Bob Garcia (chairman of the subcommittee) presiding.

Mr. Garcia. Good morning and welcome to our hearing on the census questionnaire and automation. This hearing is a continuation of our series of hearings on the Census Bureau's plans and activities on the 1990 decennial census.

Today, our focus is on the suitability of the census questionnaire content and design to the available technology. We will cover two areas:

First, we will explore the possibilities of creating a shorter census questionnaire form. We have learned that not all data collected from 100 percent of the population are processed and disseminated. In 1987 the Census Bureau is due to report to Congress on the types of questions it plans to ask on the decennial, and in 1988 the Bureau will report on the actual content of the questionnaires. In order to curtail the huge costs involved with taking a decennial, maybe there are some questions that do not have to be asked of the 100 percent of our population but to only a segment of the population.

The second area we will look into is the data conversion methods which the Census Bureau has considered for the 1990 decennial. By data conversion methods, I mean the ways in which the Bureau will convert data from questionnaire forms to computers. I want to find out the reasons behind the Bureau's recent decisions to rely on technology that dates back to the 1950's.

Through this hearing, I would like to find out what obstacles the Bureau confronted that may not have allowed them to take full advantage of modern advanced technology. Considering that we are only 4 years away from the 1990 census and that the Bureau has made its decisions regarding automation as it relates to the questionnaire content and design, it looks as though the 1990 census is already a lost opportunity. But perhaps in carefully reviewing the background of the Bureau's decisions, we can look into opportunities for the decennial in the year 2000.

I would like those who are here representing the Bureau of the Census to know that yesterday we had a meeting with GAO. We went over some of the problems that they anticipate. I want to be perfectly fair in terms of letting you know that these discussions have taken place. There are some points that they brought up in yesterday's meeting with me and the staff of the Subcommittee on Census and Population that I hope we would be able to get to today.

Sometime after you finish your presentations, I would like GAO to come up. Then we can have an open dialog as to what the views are and how, hopefully, GAO, as the body monitoring what Census is doing, can be helpful in a constructive fashion to those of you at the Bureau of the Census who have to make these decisions.

I just want to make it very clear—we are here as a committee offering absolute cooperation trying to find out how the three of us—that is GAO, the Bureau of the Census, and Congress—can work actively together to ensure that we get a census that will be cost efficient and accurate. I think that is really going to be the key to any dialog we have today.

I would go a step further. I would hope that we would be able to conclude this hearing within 1 hour 15 minutes, 1½ hours tops. There may be some votes that come up on the floor of the House at approximately 11:15 to 11:30 a.m. So what we will try to do is to expedite this.

Now, Ms. Susan Miskura, who is the Chief of Decennial Planning Division, U.S. Bureau of the Census; Gene Dodaro, Associate Director, General Government Division of the U.S. General Accounting Office; and Gail Franke, who is the vice president of Federal Government marketing, National Computer Systems, are our witnesses today.

OK, what we will do is start with Ms. Miskura.

STATEMENT OF SUSAN MISKURA, CHIEF OF DECENNIAL PLANNING DIVISION, U.S. BUREAU OF THE CENSUS

Ms. **Miskura**. Thank you. Mr. Chairman, I will make brief remarks here and submit my entire testimony for the record.

With me is Peter Bounpane, Assistant Director for Demographic Censuses.

I will discuss four topics today related to census planning.

The first topic I will talk about is the content of the questionnaires for the 1990 Census of Population and Housing. As we seek to determine questionnaire content for the census, we have one overriding goal: We must balance the needs for data against the length of the census questionnaire and the amount of time it takes respondents to fill it out.

On the one hand, we must make sure that the 1990 Census of Population and Housing collects all the critical data our Nation needs to address population and housing issues in the 1990's and beyond.

On the other hand, we realize that public cooperation could be undermined if the questionnaire is too lengthy, or contains items that do not meet important public needs.

We believe that we struck the proper balance for the 1980 census. Public cooperation and acceptance of the importance of the census was excellent. In 1980, over 80 percent of the households returned the census questionnaire by mail. This is quite an achievement in a society as complex and mobile as ours.

To make sure that we ask only those questions that meet important public needs, we have held discussions with data users in a number of forums. These include local public meetings, special conferences, interagency working groups, and the Federal Agency Council. During these discussions, we are hearing many more legitimate and valid data needs than we can possibly satisfy.

The census must collect data that are required to meet demonstrated public needs, or that are required to fulfill legal mandates and implement governmental programs. We asked the Federal agencies to identify all legal mandates and programs requiring certain data.

Census Bureau specialists apply a number of other criteria to determine a set of potential items for inclusion on the questionnaire. We then test proposed new items, or modified wording, format, and sequencing for items that have been asked in previous censuses.

The National Content Test is our main testing vehicle. It is designed to provide information on the reliability of the data collected, and the ability and willingness of respondents to answer the questionnaires.

I mention in our written testimony some of the new items we are testing. At this time, Mr. Chairman, we believe that the questionnaires, both the long and the short form, for the 1990 Census on Population and Housing will be about the same length as they were in 1980.

We plan to ask the same population questions on a 100-percent basis as in 1980. The population questions asked of a sample of persons probably will be similar to 1980, but we are awaiting the results of the National Content Test before making a final determination.

Three questions that appear on the housing page are really coverage questions. We are testing the possibility of reducing these to one question for 1990. We expect most of the housing questions asked on a 100-percent basis in 1980 will remain. We do plan to move the question on complete plumbing facilities to the sample questionnaire.

The 100 percent housing questions were supported for inclusion by the interagency working group on housing issues. They also received strong support from local planners, especially from urban centers, who are major users of the data for census blocks within their cities. Several of the 100 percent items define housing units and help us to insure complete coverage of the population and housing inventory.

With regard to the sample housing questions, we plan at this time to eliminate those dealing with stories and structure, elevator and structure, and cooking fuel. We are considering having one question instead of two on the number of automobiles and trucks in the household.

We are testing the possibility of collecting certain housing data common to all units in a multiunit building by means of a structure questionnaire. This questionnaire would be administered to a knowledgeable respondent such as an owner, manager, or superintendent of the building. This approach might enable us to collect more accurate data and reduce the number of questions asked of households in multiunit structures.

We will report to our oversight subcommittees as the result of our tests of questionnaire content become available.

The second topic I will discuss, Mr. Chairman, is questionnaire design research, including focus groups. It is important to design a census questionnaire that enhances the response to the census, and the correct completion of every item on the questionnaire. That is why we have given much attention to the subject in the past, and why we are currently doing so as we plan the 1990 census.

We conducted a number of studies in connection with the 1980 census and are conducting studies related to questionnaire design in our 1986 test censuses and in the National Content Test.

One of the things we have found in our studies is that the entire questionnaire package bears looking at. That is, the envelope and the inserts as well as the questionnaires. Thus, in the National Content Test this year, we are testing two different envelope designs to see their effects on the mail return rates, response rates for individual questions, and data quality. One envelope was designed to be attractive, the other to look official.

In the 1986 test censuses in central Los Angeles County and east-central Mississippi we are testing the effects of including a motivational insert in the questionnaire mailing package. This test is designed to see whether a brief written appeal for cooperation accompanying mailing out of the census form can improve mail response rates and reduce question nonresponse.

Some of our studies have also examined the effect of having our questionnaire designed to be read by FOSDIC. We use FOSDIC to convert questionnaire data to computer-readable format in the 1960, 1970, and 1980 censuses. FOSDIC does impose some constraints on questionnaire design in terms of how questions are formatted. Most of the questions are designed to be read by filling circles.

We conducted an alternative questionnaire experiment as part of the 1980 census in which we compared the 1980 FOSDIC questionnaire to an alternative FOSDIC questionnaire and to a questionnaire that was a non-FOSDIC form. There was very little difference in the mail return rates for the three questionnaire versions. We are now hiring a contractor to do additional research on questionnaire design.

Focus groups which typically consist of a dozen or so participants recruited from a target population are one medium for observing reactions to questionnaires. They can be used to gather ideas for more systematic research. We conducted focus groups as part of the 1985 test census in Tampa, FL to gauge reaction to the optical mark recognition questionnaire. Participants reacted reasonably well to overall design of the questionnaire, but they mentioned problems with some questions due to format constraints. Prior to both the 1986 test censuses in both Los Angeles and Mississippi we conducted focus groups as part of our Outreach Research Program. In both sites, focus group participants raised questions about why we were doing the census in general, and why we asked specific questions. Questions were raised about several population items, including race and Spanish origin; about the coverage questions; and about some of the housing items. Many of the comments indicated that there is a need for more outreach and education in the decennial census: How it benefits society and the fact that it is both mandatory and confidential.

We also conducted focus groups after the mailout of questionnaires in Los Angeles as part of our research into nonresponse. Preliminary observations from Census Bureau observers are that few participants mentioned questionnaire design or content as reasons for failing to return the questionnaires.

The third topic I will discuss is the technology we will use in 1990 to convert questionnaire data into computer-readable format. As we mentioned at the May 1 hearing before this subcommittee, we recently decided that FOSDIC will be our primary data-conversion technology for the 1990 census.

With the FOSDIC system we film the questionnaires and use FOSDIC machines to read the data into the computer. We plan to set up FOSDIC systems in 10 to 14 processing offices. FOSDIC is a fast and accurate technology that has worked well for us in recent censuses. Of course, we will upgrade FOSDIC to make it even faster and more accurate in 1990.

We will use keying as a supplement to FOSDIC for entering some of the handwritten data on the questionnaires into computer-readable form.

The fourth topic I will discuss is optical mark reading technology. We have considered optical mark recognition as an alternative to FOSDIC early in our planning and we tested it in our 1985 test census in Tampa, FL.

One of the primary reasons we considered OMR for data conversion was that it might give us more flexibility in decentralizing the census processing system. I will discuss the issues we considered in reaching the decision not to use optical mark recognition technology as the primary data conversion technology for 1990.

First of all, there were technological concerns. The standards that must be met by any technology for converting decennial census are high. The OMR scanner we tested did not meet the proven speed and accuracy of the FOSDIC system that we used in 1980.

Second, there was the issue of decentralization. During this test we discovered that the technology required a carefully controlled environment, in terms of temperature and humidity. These requirements would place significant limits on the type of space that we could obtain for the more than 400 collection offices in which we hope to use OMR. Based on this experience, we were concerned about the requirements for maintaining a widely distributed OMR system.

Third, there was the issue of timing. We did not believe that a reliable OMR scanner that corrected the problems we had observed could be fully tested before our automation decision date of September 1986.

Finally, there was the issue of cost. We estimated that there would be substantial developmental costs in design and fabrication of an OMR scanning system that would meet our needs. Our decision was based on available evidence from the 1985 OMR test experience, the current state of the technology, the potential for decentralization, the census equipment acquisition schedule, and the costs involved. However, we do expect to consider OMR as part of a research and development program for the year 2000 census.

Mr. Chairman, that concludes my testimony on questionnaire content, design and processing. I would be willing to answer any questions you might have.

Thank you.

[The statement of Ms. Miskura and her response to written questions follow:]

STATEMENT OF THE CHIEF OF THE DECENNIAL PLANNING DIVISION OF THE BUREAU OF THE CENSUS

Susan M. Miskura

Before the Subcommitte on Census and Population
Post Office and Civil Service Committee
U. S. House of Representatives
May 15, 1986

Mr. Chairman, thank you for this opportunity to brief the Subcommittee further on plans for the 1990 Census of Population and Housing. As you requested, I will discuss four topics related to census planning: (1) questionnaire content, (2) questionnaire design research and focus groups, (3) data conversion methodology, and (4) optical mark recognition technology.

Questionnaire Content

Mr. Chairman, the first topic I will talk about is the content of the questionnaires we will use in the 1990 Census of Population and Housing. We also discussed this topic at the hearing before this Subcommittee on September 26, 1985.

The decennial census is the Nation's primary source of data for small geographic areas and small population groups. A general principle governs the selection of subject content for the census: The census must be aimed solely at data that are required to meet well demonstrated public needs or that are required to fulfill legal mandates or implement governmental programs.

The 1990 Census of Population and Housing will mark the bicentennial of census-taking in our country. From the very first enumeration in 1790, the census has always been more than a simple headcount of the population. It has asked questions that mirror the concerns of our society. Over the decades, as our society became more complex and our government more sophisticated, we added questions to the census to meet new needs. By 1900, the census covered most of the population questions we ask today. Concerns about the Nation's housing during the Depression in the 1930's led to the addition of a set of housing questions in 1940. The Census of Population became the Census of Population and Housing.

As we seek to determine the questionnaire content for the 1990 Census of Population and Housing, we have one overriding goal: We want to balance the needs for data against the length of the census questionnaire and the amount of time it takes respondents to fill it out. On the one hand, we must make sure that the 1990 Census of Population and Housing collects all the critical data our Nation needs to address population and housing issues throughout the 1990's and beyond. These data are used for many important purposes, from apportionment and redistricting to planning and implementing social and housing programs and developing economic policy. On the other hand, we realize that public cooperation could be undermined if the census questionnaire is too lengthy or contains questions that do not meet important public needs.

We believe we struck the proper balance for the 1980 census. Public cooperation and acceptance of the importance of the census was excellent. Over 80 percent of the households returned questionnaires. This is quite an achievement in a society as complex and mobile as ours, especially when we realize that there are many factors that can contribute to nonresponse in the census.

The Applied Behavior Analysis Survey (ABAS), which we conducted during the 1980 census, indicated that questionnaire length was not a significant contributor to nonresponse. The survey found that the primary contributor to nonresponse was the reported failure to receive a form in the mail. Subjective factors such as how difficult the respondent thought the form would be and how long the respondent thought it would take to complete the form were associated with whether or not the form was started. On the other hand, objective measures of respondent burden, such as type of form received (short or long) and household size were not associated with whether the form was started. Further, households receiving long forms were just as likely as households receiving short forms to start filling them out despite the fact that the long forms were perceived as more difficult to complete than short forms. This is also supported by the fact that the 1980 mail-return rates for short and long forms were not significantly different.

Still, we believe there should be no increase over 1980 in net questionnaire content for the 1990 census. We are looking for ways to shorten the questionnaires; but, as we hold discussions with a broad array of data users, we are hearing many more legitimate and valid data needs than we can reasonably satisfy. At this time, we believe the questionnaires (short form and long form) for the 1990 Census of Population and Housing will be about the same length as those used for the 1980 census. In 1980, the short-form questionnaire asked at every household contained the 7 population questions asked of each person, 9 housing questions, and 3 coverage questions asked of all the population. The coverage questions are designed to make sure we count everyone at an address who should be counted there. The long-form questionnaire contained these questions plus the additional questions asked of only a sample (about 20 percent) of the population.

To make sure that we ask only those questions that meet important public needs, we have held discussions with data users in a number of forums. Local Public Meetings (LPMs), cosponsored by the Census Bureau and local and state organizations, were primary sources of information on the uses of the data at the state and local level. The LPMs afforded a wide variety of users, from the private and public sectors alike, the opportunity to comment on the adequacy of the data and to suggest new or modified data elements for the upcoming census. At least one meeting was held in every state and we completed the last of the 65 meetings in October 1985. Other forums and special outreach efforts--such as conferences dealing with housing issues or the needs for data on race and ethnic groups--also are major sources of suggestions on the content of the 1990 Census of Population and Housing.

For determining Federal data needs, we have sought counsel from other agencies--through 10 Interagency Working Groups and through OMB's Federal Agency Council on the 1990 Census. We asked the Federal agencies to identify all legal mandates or Federal programs requiring certain data. These exchanges have been important channels of communication.

Census Bureau specialists also apply a number of other criteria to determine a set of potential items for inclusion on the questionnaire. These include, for example, whether the data are needed for small geographic areas or small, widely dispersed population subgroups. We then test proposed new items and modified wording, format, or sequencing for questions that were asked in the previous census. The testing program will help us determine which of the many valid data needs can be pursued for the census.

We have conducted several content studies during the past few years. The National Content Test, which we are conducting right now, is our main testing vehicle. This test is designed to provide information on the reliability of the data collected and the ability and willingness of respondents to answer the questions. The mailout for the National Content Test occurred in late March 1986, followup will continue through the summer, and we will complete analysis of the results this winter. This will allow us to report to Congress by April of 1987 on the proposed subjects for the 1990 census. Additional smaller-scale tests will be needed after that as we decide on final question wording.

Planning and consultation to date have identified numerous new subjects for testing. The National Content Test forms, which include both new proposals and traditional census questions, contain about twice as many inquiries as were on the 1980 census forms. Testing will help us narrow the list of candidate questions, particularly if we find there are high nonresponse rates or data quality problems with some of the questions. Some of the proposed new or expanded topic areas we are testing in the National Content Test include--

Population:

- highest educational degree held (in addition to, or as a substitute for, years of schooling)
- disability limitations for children, and limitations in self-care and mobility for the population in general
- receipt of benefits from government programs such as food stamps, medicare, medicaid, and energy assistance
- health insurance and pension coverage
- pension income
- second jobs
- vocational education

Housing:

- identification of residential care facilities
- secondary heating fuel and equipment
- identification of cooperatives
- identification of congregate living units
- presence of smoke detectors
- condominium fees
- mobile home costs

New questions can be added to the census only if we can identify 1980 questions that are no longer needed and can be removed; or by employing innovative sampling techniques that would allow us to ask more questions without increasing the reporting responsibilities for any one household. We are currently reviewing and discussing content sampling options.

Now, Mr. Chairman, I will briefly outline some of the content changes we are considering based on what we heard in our discussions with local, state, and Federal data users. At this point, we plan to ask the same population questions on a 100-percent basis as in 1980: name, relationship to householder, sex, race, age, marital status, and Spanish-origin. (The wording, format, or sequencing of these questions could change somewhat based on our test results.) The population questions asked of a sample of persons probably also will be similar to 1980, but we will await the results of the National Content Test before making a final determination.

We are testing one question that could replace the three coverage questions that were asked on a 100-percent basis in 1980. ("Did you leave anyone out of Question 1 because you were not sure the person should be listed?", "Did you list anyone in Question 1 who is away from home now?", "Is anyone visiting here who is not already listed?") Although these questions appeared in the housing section of the short and long forms, they are not housing questions. They pertain to coverage of the population. We will reduce the number of coverage questions to one if we determine that this change has no adverse impact on our effectiveness in identifying potentially missed persons.

With regard to the 100-percent housing questions, we do plan to move the question on complete plumbing facilities to the sample questionnaire. We will also replace the 1980 question on the number of units at a single address with a question on the number of units in a single structure. The latter was on the sample questionnaire in 1980.

At this time, we do not have specific plans for changes in the remaining seven 100-percent housing questions. All were supported for inclusion on the short-form by the Interagency Working Group on housing issues, which included representatives from Federal agencies that have an interest in housing programs. These questions also received strong support from local planners, including those from urban centers who are major users of the data for census blocks within their cities.

Just as we must ask a minimum of population questions to assure complete coverage of the population, several of the 100-percent housing questions--the number of units in a structure, access (separate entrance to living quarters), tenure (owned or rented), and to some extent, number of rooms--are essential to help us define housing units and assure complete coverage of the housing and the population universes. The 100-percent housing questions also serve to provide benchmark data that make it possible to use the information in the sample questions. There is also another reason for not moving some of the questions to the sample questionnaire: For some characteristics, such as whether a housing unit is part of a condominium or cooperative, the universe is too small to allow us to collect adequate data on a sample basis.

With regard to the sample housing questions, we plan, at this time, to eliminate those dealing with stories in structure, elevator in structure, and cooking fuel. We are considering having one question instead of two on number of automobiles and trucks in the household.

We are testing the possibility of collecting certain housing data common to all units in a multiunit building by means of a structure questionnaire. This questionnaire would be administered to a knowledgeable respondent such as the owner, manager, or superintendent of the building. This approach might enable us to collect more accurate data and reduce the number of questions asked of each of the households in multiunit structures. Among the questions that could be included on the structure questionnaire are units in structure, the year the structure was built, type of heating fuel, source of water, whether a housing unit is part of a condominium or a cooperative, and so forth.

Questionnaire Design Research and Focus Groups

The second topic I will discuss, Mr. Chairman, is questionnaire design research, including focus groups. It is important to design a census questionnaire that enhances both the response to the census and the correct completion of every item on the census questionnaire. That is why we have given much attention to this subject in the past and why we are currently doing so as we plan the 1990 census.

While we ask focus groups to consider certain aspects of questionnaire design, they can be used to collect insights into a number of subjects. Focus groups typically consist of 8-12 paid participants recruited from the general public according to a set of specifications that allow us to target selected population subgroups. Focus group research seeks to develop qualitative insight and directions for future research on selected issues. Focus groups can uncover underlying motivations that can't be measured with a survey instrument.

Participants respond to situations and inquiries posed by a trained moderator. Answers and reactions are probed at the option of the moderator so they might be documented for review by the sponsor. I will discuss later our experiences with focus groups in the 1985 and 1986 test censuses.

In our research, we look both at the questionnaire itself and at the entire questionnaire mailing package, including the outgoing and return envelopes and the instructions.

I will begin by reviewing the research we conducted as part of the 1980 census that relates to questionnaire design. I mentioned the Applied Behavior Analysis Survey earlier. We also conducted a Content Reinterview Study, in which we evaluated the quality of some of the data collected in the census. We evaluated the degree of response variability in cases where identical questions were asked in the reinterview and the census and the degree of response bias in cases where more detailed probing questions were asked in the reinterview than in the census. Information on data quality from the Content Reinterview Study coupled with information on item nonresponse gives us an indication of which census questions might pose problems for respondents.

We also conducted a Knowledge, Attitudes, and Practices Survey to evaluate the 1980 public information campaign. We interviewed a sample of households to enable us to evaluate the campaign particularly among minority populations. (The sample was designed to allow us to get results for the Black and Hispanic populations separately.) There were some important findings from this survey that relate to questionnaire design. The survey showed that the public information campaign appeared to stimulate cooperative mail response behavior, especially among Blacks, Hispanics, and low-income segments of the population. Still, 25 percent of the respondents reported no exposure to the census prior to receiving the census questionnaire. Some groups might benefit from motivational messages on the questionnaire or in the mailing package. I will discuss later studies we are conducting in 1986 that relate to the questionnaire mailing package.

We also conducted the Alternative Questionnaires Experiment in 1980 to test several variations in the design of the questionnaires. The standard census questionnaires were designed to be read by FOSDIC (Film Optical Sensing Device for Input to Computer), which I will describe in more detail later. We compared two alternative questionnaires (short and long-form versions of each) to the standard FOSDIC form. The first alternative questionnaire was similar to that used in the 1970 census: It was FOSDIC-readable and collected the household roster information in a linear rather than the columnar format we used in 1980. That is, person sections ran left to right across the page rather than from top to bottom. The second alternative questionnaire, designed by an outside contractor, was not FOSDIC-readable. It presented the questions in a different format, it used color differently, and it altered question wordings. One effect of these changes was that the non-FOSDIC questionnaires were longer than the standard census questionnaires. The non-FOSDIC short form was 8 pages compared to 4 for the standard short form, and the non-FOSDIC long form was 32 pages compared to 20 for the standard long form.

There was very little difference in the mail-return rates for the three questionnaire versions. The non-FOSDIC short form was returned at a slightly higher rate than the standard short form, even though the former had about twice as many pages. We are also comparing FOSDIC and non-FOSDIC questionnaires in the 1986 National Content Test to see how the public will respond to each.

In our 1986 test censuses in Central Los Angeles County and East Central Mississippi the questionnaire covers and envelopes were redesigned to be more colorful and attractive than those for the 1980 Census. We are also testing the effects of including a motivational insert in some of the questionnaire mailing packages. As I mentioned earlier, research conducted after the 1980 census indicated that for some people, the arrival of the census mailing package was the first they had heard about the census. Thus, the census mailing package itself is a public information vehicle and is a critical information source for certain population subgroups.

This test is designed to see whether a brief written appeal for cooperation accompanying mail-out of the census form can improve mail-response rates, lower question nonresponse, and increase cooperation with follow-up enumerators. We are also looking at whether this general-purpose insert has comparable effects for various population subgroups.

The insert included red, white, and blue graphics and listed six reasons "to count yourself in on the census". The text was in both English and Spanish. We expect to have preliminary results from this study by late summer.

In the 1986 National Content Test (NCT), we are also comparing mail-return rates, response rates for the individual questions, and data quality for two alternate envelope designs. The 1980 Applied Behavior Analysis Survey showed that about 13 percent of nonrespondents (2 percent of all households) never opened their census mailing packages. Another 24 percent of nonrespondents (or 4 percent of all households) opened the envelopes but did not start to answer the questionnaire. Thus, the studies we are conducting as part of the NCT are aimed at finding new ways to get more people to open the envelopes and start answering the questionnaires.

One envelope was designed to be attractive and appealing. The assumption is that an attractive appearance motivates recipients to open their envelopes and respond to the enclosed questionnaire. The second envelope was designed to capitalize on the official nature of the census and attempts to use the authority and importance of the Federal Government to motivate recipients. The "attractive" envelope has red, white, and blue graphics; the "official" envelope is simply black print on white with no graphics. We will have results from this study during the summer.

We are issuing a request for proposals for additional work on questionnaire design. We are asking the contractor to observe and report to us on respondent behavior in filling two kinds of long-form questionnaires--FOSDIC readable and key-entry. The contractor will provide suggestions for improving both kinds of questionnaires and will design and develop alternative long-form questionnaires compatible with our data conversion techniques and other forms design considerations. We expect this project to be completed by late this year.

Now I will discuss our focus group research in both the 1985 and 1986 test censuses.

We used group research to assess the public's reaction to optical mark recognition (OMR) questionnaires in the 1985 Test Census of Tampa, Florida. OMR scanning requirements place some constraints on the size and appearance of the questionnaire. For example, the scanner used in the Tampa census could only accommodate a questionnaire measuring 8 1/2 x 11 inches, which meant that all questions and answers had to be confined to this space. To learn how these design constraints affected respondents, we arranged for a contractor to conduct four focus groups comprised of middle-income Whites, low-income Whites, low-income Blacks, and low-income Hispanics who were residents of the Tampa test census area. The contractor conducted group sessions approximately two weeks after Census Day. Discussion focused on the OMR short form, which 80-percent of the households received. We did not use OMR technology to process the long-form questionnaire.

The focus group participants reacted reasonably well to the overall design of the OMR questionnaire. Nevertheless, there were several comments directed toward design elements. Participants said that questionnaire instructions were complete and understandable, although it was apparent that people seldom bothered to read or consult them. Specific recommendations included repositioning Spanish language directions from the bottom back face to a more prominent location and researching alternative formats and contents.

In summary, OMR format constraints led to some question-specific problems that needed solving; a few other problems with the questionnaire not related to the OMR format also surfaced. I will discuss OMR in more detail later.

In the 1986 test census in Los Angeles we decided to use focus groups to gather information to help determine why people responded or did not respond to the census. These focus groups were part of a larger Census Community Awareness Program. We recruited six focus groups, each consisting of 12-15 persons who were grouped based on whether they were respondents or nonrespondents, and whether they were Hispanic, Black, or Asian.

The focus groups were conducted by a private contractor and a final report is due from the contractor in May. At this time, only general observations from Census Bureau observers on the focus group sessions are available.

In general, nonresponse appeared to be related to lack of foreknowledge of the census, the interference of personal life events (father having a stroke, a child destroying the questionnaire,etc.), or lack of motivation, including misunderstanding of the purpose or intent of the census. Few members of the focus groups said that questionnaire design or content was the cause of the failure to return it.

Data Conversion Methodology

We recently decided that our primary data conversion technology for the 1990 census will be FOSDIC. We believe that this decision represents the best balance of staffing, equipment, and workload considerations as they relate to the processing and collection offices.

FOSDIC is an acronym for Film Optical Sensing Device for Input to Computer. A version of FOSDIC was first used in the 1960 census and has been our data conversion technology in each subsequent census. The complete data-conversion system consists of automated high-speed cameras that film the questionnaires, film developers to process the raw film into rolls of microfilm, and the FOSDIC machines that scan the microfilm and record data on computer tape. We call this the FACT system which stands for FOSDIC and Automated Camera Technology.

We also considered using keying as a primary data conversion technology and earlier in our planning for 1990, we evaluated the use of a third technology--optical mark recognition(OMR). The FACT system is fast and accurate and worked very well in the 1980 census. But there are technical limitations to how many FACT systems we can build and maintain for 1990. We had been considering, therefore, data keying and OMR technology to give us maximum flexibility in decentralization. I will discuss later our experiences with OMR and the reasons why we are no longer considering it for the 1990 census.

Both FOSDIC and keying are tested methodologies that have proven workable over the years. Keying poses few constraints on questionnaire design. FOSDIC does pose design constraints(form size and layout, answer circles, etc.) and technical constraints(paper quality, etc.). However, we have used FOSDIC successfully in each census since self-enumeration was first introduced in 1960. In the 1960 census, one of the criteria we used to determine whether we could extend the mail-out/mail-back census method(and, thus, self-enumeration) was the degree to which answering FOSDIC-readable questionnaires posed problems for respondents.

Mail-return rates, response rates to individual items, and data quality generally have been quite good in recent censuses. Nevertheless, we would like to make improvements and continue to investigate ways to do so. There is little evidence, however, that the questionnaire design constraints imposed by FOSDIC seriously affect any of these three areas. Still, we conducted the Alternative Questionnaires Experiment as part of the 1980 census and plan to conduct additional research in this area for the 1990 census.

If we had decided to fully decentralize the processing, we would have had to use keying as a primary data conversion methodology in some offices. Keying would not be a viable option as the sole data conversion technology for the entire census because of the large numbers of keyers and key stations that would be required. And we determined that having two primary data conversion technologies would have excessively complicated our processing system for 1990. We will use keying only as a supplement to FOSDIC for entering some of the handwritten data on the questionnaire into computer-readable form.

The issue of "how" we will convert questionnaire data into a computer-readable format is intimately tied to the issue of "where" the data conversion will take place. These two issues--"how" and "where"--are the major issues confronting us as we plan to implement concurrent processing in the 1990 census. We reported on the "how" and "where" issues and on our plans to conduct concurrent processing at the hearing before this Subcommittee on May 1. The essence of concurrent processing is that we want to begin the conversion of questionnaire data to machine-readable form concurrently with the questionnaire collection phase.

The "where" issue involves the number of processing offices and the degree of centralization or decentralization. In 1980, when we processed the census questionnaires sequentially, we had three processing centers. More processing centers are needed to make concurrent processing a feasible option, primarily because of the need to move materials quickly between processing and collection offices. Also, centralization of processing activities would require us to hire large numbers of employees in one employment area. We must weigh these concerns against problems related to decentralization, such as the need for more hardware in a greater number of locations and the difficulties of controlling and supporting many processing offices.

At the May 1 hearing, we announced our decision to have 10-14 processing centers in the 1990 census, where we would use FOSDIC to convert the data to machine-readable form. For district offices in certain high population density areas the processing centers would receive the questionnaires, perform automated check-in using laser sorters, and perform an automated review of the questionnaires (edit), as well as data conversion. The district offices will be able to concentrate on field follow up activities for households that did not mail back their questionnaires or that mailed back incomplete questionnaires. District offices in the rest of the country will receive the questionnaires, use pen-shaped wands to perform automated check-in, and conduct clerical reviews (edits). Once questionnaires pass the edit, they would be sent on a flow basis to a processing center for data conversion, using FOSDIC. Here they will also undergo a computer review for quality assurance.

Optical Mark Recognition Technology

As I mentioned earlier, the Census Bureau has decided to cease the testing and consideration of optical mark recognition (OMR) technology as a primary data conversion methodology for the 1990 decennial census. Now, I will discuss our experiences with OMR and the issues we considered in reaching the decision not to test it further.

We undertook the investigation of OMR technology with the hope that a commercially available scanner could be found with the ability to meet the unique requirements of a decennial census data processing environment. We requested and received funds for a 1984 research and development effort, and initiated a competitive procurement process to acquire an OMR device for use in the 1985 test census. The General Services Administration issued the contract order to National Computer Systems (NCS) for its W201 OMR scanner.

NCS is recognized as a leader in the area of OMR technology. We determined that among commercially available scanners, its W201 had the best potential for successful performance in the census environment. Use of this machine placed certain limitations on census operations (e.g., the size of paper we could use for the questionnaires, the type of marker the respondents had to use, etc.). Although the Census Bureau recognized these limitations to the use of the existing machine, we believed the 1985 test experience would provide important information about the effects of these limitations on census processing requirements. We were hopeful about the potential for dispersing the equipment to perhaps as many as 400 collection offices.

The OMR test in our Tampa, Florida, test census was tailored to meet the operational guidelines of the W201 scanner as defined by NCS, even though these guidelines differed greatly from what the Census Bureau would require during an actual decennial operation. For example, the scanner could only accommodate a questionnaire measuring 8 1/2 x 11 inches. The Census Bureau designed the test questionnaire within this limitation. The scanner was engineered to read No. 2 lead pencils, and so the required pencil was enclosed in every OMR mailing package sent to Tampa households. A special data keying operation was set up to capture and convert information from the sample questionnaires sent to 20 percent of the households in the test census area since the scanner could read only single sheets and could not scan the multipage census booklets. These sample multipage booklets, although used to enumerate only one-fifth of the households during a census, represent as large a decennial processing workload as the short-forms.

We carefully identified hardware, software, and operational problems with the scanner. The influence of these problems on the 1985 test and their implications for the 1990 Decennial Census were evaluated and are fully documented in a report issued in July 1985.

Mr. Chairman, I will now summarize the major findings from this report. First there were environmental concerns. We discovered during the testing and debugging of the OMR system that stringent environmental conditions had to be maintained to limit hardware problems, and questionnaire feeding and reading problems. We had to subsequently install individually controlled humidifiers, dehumidifiers, heaters, and air conditioning within the specially modified space provided to house the scanner in the Census Bureau's permanent processing facility in Jeffersonville, Indiana. The scanner was sensitive to temperature and humidity; unless the questionnaires were acclimated to the controlled environment they could not be scanned. These environmental concerns are particularly important, because they would create stringent requirements for the space for housing decentralized OMR systems in 1990.

Second, there were questionnaire design and printing concerns. The questionnaire had to be designed within a double-sided 8 1/2 x 11 inch area. All questions, instructions, and response fields were forced into this limited space. Also, paper specifications proved to be restrictive and had to be altered before printing could be completed.

Third, there were paper transport problems. To limit scanning problems encountered with the folded census forms, the vendor lowered the input hopper capacity from 500 to 150 forms; we later had to lower the capacity to 100 forms. This significantly slowed production.

Fourth, the scanner was unable to determine when more than one circle within an answer field was filled in. Detection of multiple marks is important so that only complete data would be accepted as valid.

Fifth, there were difficulties in evaluating scanner accuracy. The scanner was designed only for test scoring application where the documents to be scanned are filled out in a controlled environment. But in an uncontrolled census environment answer marks on the census questionnaires are often light, not precisely filled within the circle wall, or made by any implement other than a pencil, and multiple entries or erasures are sometimes made within an answer field. Problems occurred in the test even though we provided respondents with the proper marking pencil, something that may not be possible in a decennial census. Although we did not evaluate the scanner's reading performance under these uncontrolled conditions, a decennial census is by its nature an uncontrolled environment and we remain concerned about these problems.

In summary, Mr. Chairman, the results of all of these tests indicated that the W201 scanner could be a viable alternative for decennial census data capture only if substantial modifications to the census system were made. NCS acknowledged this fact and submitted an unsolicited proposal to develop and produce a prototype of a census-ready machine even before the first Tampa form was processed.

In early planning for the 1986 Test Census of Central Los Angeles County, we considered further testing of OMR technology. In July 1985, it became clear that a revised OMR scanner could not be produced and procurred in time for use in the 1986 test census. In September 1985, we issued a Request for Information (RFI) to determine private sector capabilities to produce a satisfactory machine in time to be evaluated thoroughly in special purpose tests before the deadline for processing decisions in September 1986. After issuing the RFI, we decided that pursuing OMR technology as the primary data collection methodology for the 1990 census was not a viable option. This decision was based on four critical issues, which I will now summarize.

First, there were technological concerns. The standards that must be met by any technology for converting decennial census data are, by necessity, high. If any new technology for data conversion is to be considered a viable alternative, it must be shown to perform favorably when compared to the system for conversion used during the 1980 census.

The OMR scanner we tested failed to meet the proven speed and accuracy of the system used in the 1980 census--FOSDIC.

Second, there was the issue of decentralization. One of the primary reasons we considered OMR for data conversion was that, like key-entry, it might give us more flexibility in the degree of decentralization of the census processing system. Any highly decentralized processing system for a decennial census, however, must have very low maintenance requirements. Based on the 1985 test experience, we were concerned about managing a widely distributed OMR system.

Third, there was the issue of timing. We face a very tight schedule for census decisionmaking about automation of data collection and processing operations because all required equipment for 1990 must be fully tested and in place for the 1988 dress rehearsal census. We had set a goal of deciding what system(s) will be used for primary data capture during the 1990 census by September 1986. Some outside observers believed even this schedule left too little time for procurement and implementation.

We did not believe that a reliable, OMR scanner that corrected the problems discovered in the 1985 test census could be tested fully before the automation decision date of September 1986. A best case scenario would be to procure and test an OMR device by September 1986 and contract for production beginning in September 1987. The required modifications to the scanner, although not impossible, represent significant engineering and technological changes that must be identified, designed, built, thoroughly tested and debugged, with final alterations made and retested before the scanner could be placed into production in preparation for use in 1990. Even if a prototype of a full-featured census-ready scanner could have been built, the manufacturing of sufficient numbers of machines to distribute OMR to many locations could not be guaranteed.

Finally, there was the issue of cost. There were substantial development costs in design and fabrication of an OMR scanning system that would meet the Census Bureau's basic technical and decentralization needs.

We also questioned whether an OMR system developed to meet unique decennial data conversion requirements would have a high remarketing potential, or be of use to other Census Bureau programs after the 1990 Decennial Census processing is completed.

Our decision was based on available evidence from the 1985 OMR test experience, the current state of the technology, the potential for successful decentralization, the census ADP equipment acquisition schedule, and the significant costs involved. However, we will consider OMR as part of a research and development program for the 2000 census, along with high-resolution image capture and other technologies. By so doing, we will further our efforts to accomplish more efficiently and accurately what is a very difficult and complex data processing operation.

CONCLUSION

Mr. Chairman, that concludes my testimony on questionnaire content, design, and processing. At this time, we expect the 1990 census short- and long-form questionnaires to be about the same length as those for the 1980 census. We plan to make some changes in the specific questions or question wording and may make additional changes based on the results of the National Content Test.

We plan to process the questionnaires using FOSDIC technology, as in 1980. FOSDIC imposes some design constraints on the questionnaires, but we are undertaking research that could lead to improvements in the design of a FOSDIC-readable questionnaire and the other components of the mailing package. Although we determined not to use OMR technology in the 1990 decennial census, we will consider OMR as part of the research program for the 2000 census.

In determining which questions to include on the census questionnaire we are very careful to include only those that meet important public needs. Our experience has been that when we make the public aware of the importance of the census and why we are asking each question, we achieve good public cooperation.

I look forward to hearing any comments or questions you might have.

Responses to Questions from
Subcommittee on Census and Population
to
Susan Miskura
Chief
Decennial Planning Division
on
the 1990 Census Questionnaires and Automation
May 15, 1986

QUESTION 1. We have just heard the GAO testify that you are not going to use the most cost-effective procedures to process the 1990 census. How do you respond to these charges?

ANSWER: Many of the things we are doing new for 1990 will increase efficiency. For example, automating the geographic support systems--TIGER; having an automated address control file and doing automated check-in; and having automated management information systems will be cost-effective and will directly respond to problems experienced in the 1980 census.

We will only invest in automation that reduces costs or that is necessary for maintaining or improving the quality of the census. While we cannot know at this time whether a specific automation decision will save money, we believe that our decisions will lead to a more efficient and accurate census. Automating census operations will allow us to replace labor-intensive and error-prone clerical operations with automated techniques that are quicker, more accurate, and more controllable.

Cost-effectiveness must be examined in terms of the entire census process. What initially may appear to be cost-effective for one particular aspect of the census may not be when all other related aspects of the census are considered. Furthermore, we must also consider the risk that new technologies might fail and the costs associated with such failure.

QUESTION 2. On May 1st you announced that concurrent processing of the census forms would only occur in those areas that are hard to enumerate. According to the GAO testimony that we have heard today, this decision means that you will use a "modified 1980 system" for most the Nation. They say they are concerned that you have "foregone the benefits that could be derived from a more automated operation." Again, quoting them, they feel that your decision "compromised its goals for automation and for the census as a whole."

Now these are some pretty weighty criticisms of your plans. How do you respond to them? What steps are you taking to make sure that the goals for the 1990 census are not compromised?

ANSWER: Our processing decision, which we discussed in hearings on May 1 and May 15, allows us to perform concurrent processing for the entire lower 48 states, not just for certain areas. (We have not determined our processing plans for Alaska, Hawaii, Puerto Rico, and the Outlying Areas.)

For all areas, we will be entering completed and accepted data into the computer on a flow basis, concurrent with field operations, and much earlier than for the 1980 census. For all areas, we will have an automated address control file and automated check-in and an automated management information system. This plan will enable us to process data far sooner than we did in 1980 and therefore release data products earlier.

The major difference between the district offices in the high population density areas and elsewhere is that we will do the initial edits by computer in the former and clerically in the latter. Even where we do the initial edits clerically, we will still do computer edits later as a quality assurance measure. To do initial edits by computer everywhere would have required a far greater number of processing centers (with much more equipment and staff) than the 10-14 we now plan.

We are planning several major automation advances over the 1980 census: (1) earlier data capture, (2) automated address control file and automated questionnaire check-in, (3) automated questionnaire edit, (4) automated management information system, and (5) the automated geographic support system (TIGER).

QUESTION 3. GAO stated in its testimony that the size of the questionnaire form influencing the response rates is particularly evident in inner city areas. According to GAO, for the 1980 census, the mail return rates for the short form was over 7 percent better than the long forms in these hard-to-enumerate areas. In addition, according to GAO, rates for the short form questionnaires have been consistently higher than for the long forms in the 1985 and 1986 pretests. However, you've mentioned in your testimony that the 1980 mail-return rates for short and long forms were not significantly different. How do you respond to that?

ANSWER: In the 1980 census, for the entire Nation, the mail-return rate for the short form was only about 1.5 percent higher than for the long form. For our centralized district office areas--which included mostly hard-to-enumerate central cities--the mail-return rate for the short form was only about 2.5 percent higher than for the long form.

In the 1985 and 1986 test censuses, there were larger differences between short- and long-form mail-return rates, about 8 to 10 percent. However, it is difficult to draw any conclusions from these figures because we were testing new procedures in these sites, these are just a few localities, and there was limited outreach available. The numbers from the 1980 census are much more useful because it was the national census complete with full-scale publicity and a much larger volume of forms were returned.

QUESTION 4a. You've mentioned that one way to add any of the new questions tested at the National Content Test is to find ways to employ innovative sampling techniques that would allow you to ask more questions without increasing the reporting responsibilities for any one household. How can that be done? Would you share with us some of the sampling techniques that you may be considering?

ANSWER: Now that we have decided our basic processing plan for the 1990 census, we are examining the cost and feasibility of using a kind of nested sampling technique. If we should adopt this technique, about 19 percent of the housing units would receive questionnaires that include the basic sample questions, those for which the data are needed for census tracts, places, and so forth. (This questionnaire would be very similar to the 1980 sample questionnaire.) An additional 1 percent of the housing units would receive a different sample questionnaire that includes different questions, those for which the data are needed only for very large geographic areas such as for states, the largest metropolitan areas, and certain large large counties. This would create two samples-a 19-percent sample and a 1-percent sample. The total sample size, 20 percent, is roughly equivalent to the sample size in 1980. Alternatively, the split could be 18 percent and 2 percent.

QUESTION 4b. In my understanding, the purpose of the National Content Test is to improve upon the questionnaire content of the census. Not to cast a dark cloud over your plans, but what will happen if the Census Bureau finds itself being unable to add questions to the 1990 census either because existing questions cannot be dropped or because no innovative sampling techniques are found to be suitable?

ANSWER: In that case, we would not ask the additional questions. We do not plan to increase the overall workload for respondents for the 1990 census. In our consultations with data users, we have heard many more legitimate requests for data than we can reasonably satisfy. Even by removing some existing questions and using different sampling techniques, we would not be able to satisfy all requests for new data.

QUESTION 4c: What other benefits will the National Content Test give the Census Bureau?

ANSWER: In addition to testing new questions in the National Content Test, we are also testing new wordings, formats, or placements (where the question is placed on the form in relation to other questions) for 1980 census questions that could improve the quality of the data. We are also testing the effects of different envelope designs on mail-response rates.

QUESTION 5. Please comment on GAO's finding that degree of literacy skills has a direct impact on responses to questionnaire forms.

ANSWER: We require only one knowledgeable respondent to provide data for a complete household. We realize that some persons and, perhaps even some complete households, do not possess the language skills to properly fill a census questionnaire or other types of documents. We provide telephone and walk-in assistance in filling questionnaires and provide questionnaires and instruction guides in languages other than English. If a householder cannot complete a questionnaire using these services, we complete the interview by personal visit. We send census enumerators to follow up on housing units that do not return questionnaires and telephone or personally visit housing units for which additional data are needed.

We keep the size of the questionnaire reasonable by asking only required data, and we do extensive testing on questionnaire wording, design, and so on.

QUESTION 6. At our May 1st hearing, the Census Bureau announced its decision to do concurrent processing in only hard-to-enumerate areas. Because undercount is a serious problem in these hard-to-enumerate areas, the Census Bureau developed local review programs in which local officials review the accuracy of numbers of areas covered by the census. It seems that concurrent processing may hinder the effectiveness of the local review programs. What is your response that?

ANSWER: We are committed to a successful Local Review Program for the 1990 census. Concurrent processing will not hinder the effectiveness of the program. It should improve the program by making the data provided to the local officials more accurate and complete. By converting the questionnaire data to machine-readable format earlier in the census process, we will have more time for review and correction of the data and for early identification and correction of coverage and other enumeration problems.

QUESTION 7. What are the reasons for not doing concurrent processing for the entire nation?

ANSWER: We will be doing concurrent processing for the lower 48 states and the District of Columbia. We have not determined yet our processing approach for Alaska, Hawaii, Puerto Rico, and the Outlying Areas, but we are committed to meeting the same goals for these areas--timeliness, accuracy, and efficiency--as for the rest of the country.

QUESTION 8. You've mentioned in your testimony the issue of cost related to adopting the OMR technology to the decennial processing. How much do you estimate the cost would be? What are the cost factors involved?

ANSWER: We have estimated that the research and development programs required to procure a prototype optical mark recognition (OMR) scanner to test in 1986 would cost at least $1,000,000. We also estimate that the unit cost for a production scanner for the 1990 Decennial Census would be about $150,000 each. Based on these estimates, we believe OMR technology would be more expensive than FOSDIC (Film Optical Sensing Device for Input to Computer).

Mr. **Garcia**. I thank you, Ms. Miskura.

Now I would like to invite Mr. Dodaro, who is the Associate Director, General Government Division, U.S. General Accounting Office to come on up.

As I said earlier, members of the subcommittee and I met with Mr. Dodaro for the purpose of going over what we would be talking about here today.

Mr. Dodaro, it is good to see you and I guess you know the procedure. What we would like to do is to have you submit your testimony for the record and that will be accepted without objection. If you would be kind enough to proceed.

STATEMENT OF GENE DODARO, ASSOCIATE DIRECTOR, GENERAL GOVERNMENT DIVISION, U.S. GENERAL ACCOUNTING OFFICE

Mr. **Dodaro**. Thank you, Mr. Chairman. Good morning.

On my immediate left is Jerry Donoghue, who is responsible for our work on the questionnaire, and on my far left is Johnnie Butts, who is responsible for our work on census automation.

[Mr. Dodaro presents slide presentation.]

Mr. **Dodaro**. I would like to begin with a brief overview of what we reviewed during the course of our work.

The first area we looked at was the census questionnaire itself. The design and content of the questionnaire drives many of the other decisions and activities that occur during the conduct of the census and greatly influences the outcome. The response rates, respondent burden, quality of the information, as well as the data processing options all emanate in part from the decision on the questionnaire.

The second area we looked at was the Bureau's planning activities for automating the 1990 census. Both the automation area, as well as the questionnaire area, have a major bearing on controlling the costs and improving the quality of the information during the upcoming census.

About 70 percent of the $1.1 billion spent to take the 1980 census went for data collection, preparation, and processing. Additionally, during 1980 many labor-intensive activities were part of the census operation.

In our opinion, and in the view of the Bureau, some of these areas would be fertile grounds for automating.

For example: 37,000 clerks were used in 1980 to manually edit the questionnaires. Computerized editing of the questionnaires could not only reduce the number of people required, but also would achieve greater consistency in the edit, thus improving the quality of the data. Additionally, earlier capture of the information from the questionnaires would allow greater time for local review.

There are three major areas that we would like to stress this morning.

The first is that we think the Bureau is missing an opportunity to test a more user-friendly, easier-to-complete short-form questionnaire. We think that this could increase mail response rates, particularly in hard-to-enumerate areas.

Second, we think the Bureau has not moved as aggressively as it needs to in embracing new technologies for capturing the information from the questionnaire. The Bureau started too late in its planning activities and has been moving too slowly, and now finds itself in a position of falling back upon data capture technologies that it has used in the past, because too little time is available for exploring additional options.

Third, we are skeptical that the Bureau will be able to achieve its goal of holding the cost, on a per-household basis, exclusive of inflation, in 1990 similar to that that was experienced in 1980. All indications are that the 1990 census costs will increase.

I would like to talk about each of these areas in a bit more detail.

First, the short-form questionnaire. We focused most of our effort on the short-form questionnaire, which is sent to 80 percent of the households. We looked at the justifications that were provided to the Bureau for the need for that information, particularly the housing information. We also talked to Federal agencies as to how they would use the information that is collected.

As you know, the Census short-form questionnaire contains seven population questions, three questions designed to improve the count, and eight questions on housing. We think most of the housing questions add to the length and complexity of the questionnaire, and have a tendency to decrease the response.

We think questions such as the value of property, amount of rent that is paid, as well as questions on the number of rooms, have a tendency to also detract from the Bureau's ability to obtain a quick and accurate population count earlier in the process.

A number of the focus groups, which the Bureau discussed this morning, also echoed similar concerns. They raised questions about the complexity of the form, the need and the legitimacy of the Government to ask many of the housing questions when the primary purpose of the census was to obtain the population count.

Also, during public hearings, several persons raised concern about the length and complexity of the questionnaires.

In our discussions with Federal agencies we found that many used sample data from the questionnaires even though 100 percent information was available. Also we found that Federal agencies asked for data at 100 percent for geographic levels that are already estimated from the long-form questionnaire.

We also found that some write-in information that was asked for as part of the census was never captured and used as part of the processing activities for 1980.

While we have not proved conclusively that the housing questions on the short form are not needed, all indications point to the feasibility of removing those questions and testing a more streamlined short-form questionnaire. This ought to be explored; particularly in view of the respondent burden associated with those questions—of asking them of 100 percent of the Nation's households—as well as the costs that are likely to increase as part of the 1990 operations.

What do we think the test of a short form could result in? No one knows until the test is conducted whether or not the streamlined form will increase the mail response rate. That is one of the reasons we advocate the test.

However, traditionally, the short-form questionnaire has received a higher response rate than the long-form questionnaire. While there was only a 2-percent differential nationwide, that differential was higher in hard-to-enumerate areas.

Additionally, during the 1985–86 pretests the short-form questionnaire has been receiving anywhere from 6 to 10 percent higher response rates than the long-form questionnaire.

The response rate increase is very important, since the Bureau has estimated that for every 1 percent increase in the response rate, it can save $6 million in associated costs. This would include reducing followup costs; moreover more completely filled out questionnaires would reduce some of the costs associated with sending enumerators to follow up, which is particularly occurring in the hard-to-enumerate areas. Streamlining the questionnaires would also have a tendency to eliminate some enumerator bias and we feel reduce the respondent burden.

Removing the housing questions from the short-form questionnaire would result in about 85 million households in 1990 being asked eight fewer questions, and could reduce the time it takes to complete the short form by one-third.

Now in addition to the response and cost and burden, there is also another avenue that could be opened up to the Bureau if it decides to go with the more streamlined short form. That is, in increasing its options for data capture. This is particularly important as it relates to the use of the OMR equipment, as I will talk about in a minute.

I would like to shift now from discussing the questionnaire to the Bureau's planning activities to automate the 1990 census.

We feel that the Bureau could have moved sooner to begin planning. The lack of detailed planning and advanced decisionmaking in this area has limited the Bureau's options for exploring new data technology for 1990.

In 1982 we urged the Bureau to begin seeking out new ways to automate the 1990 census. But we noted at that time that its original planning activities were not very well coordinated. The Bureau agreed and affirmed its commitment to seek out options for automating the census.

But as late as December 1984, the Secretary of Commerce, in his report, under the Federal Managers Financial Integrity Act, noted that the lack of a master plan for conducting the 1990 census was a material weakness in the Department.

The Bureau published a plan in February 1985 which set out a series of milestone decisions, but in our view the plan was not as integrated as it needed to be.

A good illustration of what has occurred as a result of not making earlier decisions and preparing detailed plans early on for the 1990 census, is the Bureau's efforts to use the optical mark reader technology.

Early on, the Bureau began exploring the use of OMR as its primary data capture technology for 1990. However, it limited its consideration of the optical mark reader to commercially available, off-the-shelf equipment. This equipment provided several potential advantages and disadvantages.

On the advantages side, it offered a one-step processing activity, as compared to the Bureau's three-step filming, developing, and scanning process used under the FOSDIC system. It also offered the advantage of being used in a more decentralized operation and easier to operate than some of the FOSDIC operations which require some knowledge in such subjects as chemistry.

On the disadvantages side, however, there were a number of limitations of the commercially available equipment to meet the Bureau's requirements.

For example, the optical mark reader would only read an 8½ by 11 inch form, where the Bureau's short form alone was approximately 11 by 28 inches.

Second, the optical mark reader could not read the Bureau's long form multipage booklet unless the pages were separated.

Third, the optical mark reader was meant to be used primarily in controlled environmental situations where people use specialized writing instruments and humidity and other environmental factors could be controlled.

Despite the knowledge that this piece of equipment could not meet the Bureau's requirements, the Bureau went ahead and tested it during the 1985 pretest. While the pretest showed the magnitude of the problems, it proved very little other than that the equipment could not meet the Bureau's requirements.

The vendor had proposed to the Bureau in January 1985 and again in April 1985 that it could develop a modified optical mark reader to more appropriately satisfy the Bureau's requirements.

The Bureau moved slowly however, on this request, and in November 1985 notified the vendor that the OMR would no longer be considered for the primary data capture in the 1990 census. As a result, the optical mark reader was not tested during the 1986 pretest, although a stated objective of those pretests was to consider alternative data capture technologies.

In summary, from the beginning the Bureau knew that the commercially available equipment would not meet its known requirements and expressed a real reluctance to change those requirements.

In our view, moving ahead and testing it, absent the possibility of changing the requirements, was not a prudent use of time. A decision could have also been made earlier in the decade to finalize those requirements and begin a research and development effort to come up with a modified optical mark reader that could have satisfied the Bureau's requirements.

Where do we stand today? As the Bureau mentioned this morning, it has made its decision to use FOSDIC as the primary data capture system for 1990. We are encouraged that the Bureau made the decision in April, which is 5 months earlier than it originally planned. And we are also pleased to see that the Bureau has not opted to use data keying as a primary data capture.

However, there are a number of concerns that emanate from this decision. Essentially, the Bureau is back to using the data capture technology that it has used since the 1960's. They have proposed some modifications. However, the modifications have not yet been fully explored, and we are likely to have many manual operations that occurred in 1980 repeated in 1990. For example: About 93 million questionnaires during the 1990 census could be manually edited again. Also, I know the Bureau is attempting to sort through some of the logistical and management problems stemming from using 10 to 14 processing offices as opposed to the 3 used in 1980.

The questionnaire, as the Bureau indicated this morning, basically will remain unchanged from the 1980 questionnaire.

There are a few other concerns that we have, that we would like to talk about this morning.

The first is that we think the cost will continue to increase for the 1990 Census. The many manual operations likely to be repeated, combined with an estimated workload increase of an additional 18 million households that the Bureau would have to canvas in 1990, all are likely to escalate the cost.

Census Bureau temporary labor costs have been reported to be increasing, in excess of the general rate of inflation and the Bureau is likely to use in the neighborhood of several hundred thousand people again as it did in 1980 to take the 1990 Census.

Additionally, as I mentioned, there is a number of potentially greater management problems that the Bureau will have to confront in using more processing offices.

What can be done at this point? The Bureau has been reluctant to test the shorter, more streamlined shortform questionnaire without the housing questions. We think such a test may be justified, although it would have to be carefully designed and explored; and the results of that test weighed against meeting the needs of the data users.

There also needs to be more attention given to coming up with costs associated with the Bureau's recent decision on processing and district office configurations. We believe cost estimates should be shared with the Congress and a dialog continued as to what are some options to reduce those costs in 1990, at least control them within a manageable level.

We also strongly urge the Bureau, as it has indicated, to continue considering alternatives including automating some of the manual activities for the 1990 Census.

As I mentioned this morning, time is running out with regard to making any sweeping changes for the 1990 Census in the area of automation, but we would urge more attention being given to computerized editing of the information.

That concludes our summary statement, Mr. Chairman. We would be glad to answer any questions at this point.

[Statement of Mr. Dodaro and his response to written questions follow:]

STATEMENT OF

GENE L. DODARO

ASSOCIATE DIRECTOR, GENERAL GOVERNMENT DIVISION

UNITED STATES GENERAL ACCOUNTING OFFICE

Mr. Chairman and Members of the Subcommittee:

I am pleased to appear today to discuss preparations for the 1990 Decennial Census in two interrelated areas, questionnaire development and data capture technology. Improvements in these areas could greatly contribute to controlling costs and enhancing the quality of the census. With me is Jerry Donoghue, who is responsible for our work on the questionnaire, and Johnnie Butts, who is responsible for our census automation work.

The Bureau's planning and preparations, including its tests and decisions to date, have led us to believe that the 1990 census will not be as cost efficient as it could be. The Bureau's reluctance to test a shorter short form, its questionable approach to procuring optical mark reader (OMR) equipment, and the questions raised by its recent decision on data capture and processing office configuration all point to missed opportunities to significantly improve upon the 1980 census.

ADVANTAGES OF A SHORTER SHORT FORM

As we mentioned in our previous testimonies before your subcommittee in June 1984 and April and July 1985, we have reservations about the size and content of the short form sent to about 81 percent of the households in 1980. We continue to believe that the short form should be limited to the basic questions needed to obtain an accurate population count--that is, questions oriented towards population characteristics and those used to improve the count. Housing questions--such as plumbing, value, and rent of housing units--increase the questionnaire's complexity and consequently discourage response. We believe that housing data obtained from the long form (sample questionnaire) may meet federal needs. This sample form contained not only the short form questions but more detailed questions on population and housing as well. One housing unit in six was asked to complete the long form in 1980 except for communities under 2500 people, where one half of the housing units were sampled.

Our more recent work has raised questions regarding the federal need for housing data from 100 percent of the nation's households. For example, we found that some federal data users were actually using sample data even though 100 percent data or data collected from all of the households was available, and that some users had requested 100 percent housing data for geographical levels for which data are also estimated from sample questionnaires. The need for housing data from all households should be more closely weighed against associated collection costs and respondent burden.

The content and design of the questionnaire is a major factor affecting response rates, quality of response, response burden, and data processing requirements. Since about 70 percent of the $1.1 billion in 1980 Decennial Census costs were incurred with the collection, preparation, and processing of the data, efforts to reduce the short form questionnaire, both in size and questions, could be cost beneficial. This is particularly important considering that the 1990 Decennial Census will record information from about 106 million households, 18 million more than for 1980. If the Bureau does not streamline and simplify the questionnaire form, it will be missing out on a potential cost savings opportunity, as well as a chance to improve the perceptions and receptiveness of the U.S. public to the census.

Cost could be reduced

The mail response rates in 1990 will have a direct impact on the nonresponse followup costs. Streamlining and simplifying the short form should improve the mail response rates for 1990, and might greatly reduce the high cost of sending out enumerators to followup with nonrespondents, particularly if better response rates are obtained from the hard-to-enumerate areas. For 1990, the Bureau estimates that each 1- percent increase in the mail response rate would save about $6 million in followup costs.

That the size of the questionnaire form influences the response rates is particularly evident in inner city areas. For the 1980 census, the mail return rates for the short form was over 7 percent better than the long forms in these hard-to-enumerate areas. Also, rates for the short form questionnaires have been consistently higher than for the long forms in the 1985 and 1986 Pretests, as shown below.

Pretest	Response Rate	
	Short form	Long form
Jersey City	39	31
Tampa	58	48
Los Angeles	36	28
East Central Mississippi	65	59

We do not know the extent to which a shorter short form will increase response rates. However, without a test no one will know. Considering that a short form will be sent to about 85 million households in 1990, and the pressures to hold down government spending, evaluation of a revised form seems worthwhile.

Some enumerator visits also could be eliminated with a shorter form because fewer respondents may return incomplete questionnaires which require enumerators to collect the missing data. Of the 64 million questionnaires that were returned by mail in the 1980 census, 13 percent of the short forms did not meet the Bureau's standards for completeness. On the other hand, 36 percent of the long forms did not meet these criteria.

A shorter, simplified short form questionnaire also could be processed at less cost. However, the Bureau must act on the size of the form to potentially benefit from data processing savings. Also, a reduced short form may provide the Bureau with cost-saving options for its data automation decisions.

Respondent burden reduced with fewer questions

Reducing the size of the 1980 short form will benefit respondents because the burden of completing the form will be reduced in terms of number of questions, time, and in the perception of difficulty.

With a short form oriented toward population and coverage only, about 85 million households would answer fewer questions. This could result in up to a one-third reduction in the time to complete the questionnaire.—

The perception of the questionnaire is an important factor. It is a burden for the respondent if the individual perceives the form to be a difficult task, and this could affect the completion of the form. The Bureau has indications that perception of difficulty is a factor affecting form completion.

For example, a Bureau post-1980 Census study found that nonrespondents attributed not starting to fill out the form to the perceived difficulty of the task. The study showed that "the easier to fill the form was perceived, the more likely it was to be started." The Bureau also found that, for the critical phase of finishing the form, nonrespondents attributed not completing the form to the "amount of work involved." In another study, the Bureau discovered that the difficulty--experienced or perceived--associated with completing a self-enumeration form can adversely affect response to subsequent items on the form.

Confirming these study findings, enumerator supervisors at the Jersey City Pretest site told us that some respondents commented after enumeration interviews that the short form appeared complicated. This concern, along with the perception that the form was too long, was also expressed by focus groups at the Los Angeles and Mississippi test sites.

PARTICIPANTS IN RECENT FOCUS-GROUP STUDIES WERE CONCERNED OVER SHORT FORM LENGTH AND COMPLEXITY

The Bureau's recently conducted "focus group" studies at the Tampa, Los Angeles, and Mississippi pretest sites showed that lower income groups generally had questions on, or objections to, the housing questions on the short form questionnaire. In addition, concerns were expressed that the short form questionnaire was too long and complex at the Los Angeles and Mississippi sites.

Focus group studies were generally done to ascertain the motivational messages and themes that would encourage hard-to-enumerate populations to respond to the census and return their forms. The focus group approach attempts to develop qualitative insight and direction for further research. While these three focus groups studies were limited to a total of 95 participants and the results cannot be projected, they provide useful indications of individuals' views of the short form questionnaire.

All three studies showed the participants had concern over the housing questions. In a Bureau observation memorandum on the Tampa study, the plumbing question caused groups to discuss why any of the housing questions were needed for the census. In the report on the Tampa focus groups, views were reported that some of the information requested for the census--e.g., value of an individual's home--was "none of the government's business." At Mississippi, housing questions such as those concerning entrance to living quarters and value of home were considered too personal, and participants wanted to know "why do they need to know all those other questions" when the census is a count of population. At the Los Angeles site, housing questions on "entrance to living quarters," "value of property" and "rent" were objected to the most.

In a memorandum of observation on the Los Angeles study, the Bureau noted that those queried believed the questions on plumbing (which were thought to be "ridiculous"), value of property and rent were considered to be too personal. According to the Los Angeles report, most of the objectionable questions related to housing items. These items were "specifically perceived to be a means of tricking people into exposing additional unreported individuals who occupy the household." Also, the number of rooms question was considered by some to be outside the purpose of a head count. One of the Los Angeles report's conclusions states that "objections were raised to some of the questions contained in the Census short form (particularly the housing-related questions) to the effect that their perceived true purpose is surreptitious."

At the Los Angeles and Mississippi sites, the short form's length and complexity concerned the participants. In Los Angeles, initial reactions to the short form were "it's much too much, too long"; "it looks too big. What they're asking, you should be able to put it on a 3 x 5 index card." The Bureau's observation memorandum on Los Angeles noted that a first reaction to the short form was "wow ... too big ... overwhelming." The Bureau's observation memorandum on Mississippi noted that people said the "form looked complicated, too long ... too nosey." The memorandum further noted that people felt that it was not very complicated when they were walked through the questionnaire. One representative quotation in the Mississippi consultant's report was "too much work to filling that out."

IMPROVED DATA ACCURACY

The degree of literacy skills affects data accuracy, and form completion. Respondents with marginal literacy skills should be better able to respond correctly to a simplified and reduced short form questionnaire. More people could respond, resulting in a more precise population count; and since less data would be collected through door-to-door enumeration, enumerator bias would be reduced.

The Bureau has not conducted any literacy tests since the 1970's when experiments were conducted to measure respondent literacy skills or the reading level required to complete the short form questionnaire. Literacy level is an important issue in developing a successful questionnaire. As such, the Bureau should have gained some insights into problems people have in understanding census questions by conducting studies with pretest respondents. While realizing that the impact of marginal literacy skills on accurate completion of the short form is unknown, it seems reasonable to assume that a simpler, reduced size, easier-to-read form could only help.

Moreover, since the collection and processing of data could be accomplished more efficiently and effectively using a streamlined short form, the population counts could be tabulated earlier allowing more time to review the results. Additionally, a streamlined short form would allow the Bureau more flexibility in considering data capture equipment.

Now, I will discuss the subject of data capture equipment.

BUREAU'S DECISION NOT TO USE OMR

We believe the Bureau's decision to discontinue consideration of OMR equipment was influenced by its late start in detailed planning, reluctance to revise the questionnaire form, and a slow procurement process. Whether OMR equipment could have been adapted for use during the 1990 census may never be known. However, because of its actions, the Bureau has excluded an option for using new technology without fully exploring its potential.

The commercially available OMR equipment considered by the Bureau had both advantages and disadvantages for a census. The OMR equipment employs a one-step process to read and record the data. The Bureau's traditional film to tape data capture system called FACT required three sequential processes--filming, film development and scanning the film. Also, on the basis of our observation of the 1985 pretest, the OMR equipment was easy to operate, and training time is short. For some of the FACT processes, a knowledge of chemistry and other technical subjects are needed.

On the other hand, the commercially available OMR was designed to process a single page form much smaller than the census short form questionnaire. In addition, the Bureau for the past several censuses has used a multipage booklet form for its long form questionnaire. The OMR was also designed for use in a controlled environment such as grading test answer sheets where the students are provided with #2 pencils. The Bureau's FACT equipment is generally not affected by these constraints.

Despite these known disadvantages, the Bureau decided to use existing OMR equipment in its 1985 pretest, and as a result the pretest proved very little. Most of the major problems that occurred with the OMR in the test had previously been known. The test simply helped to identify the magnitude of these problems.

The prospective vendor had made several proposals to overcome the limitations by designing a modified OMR but the procurement effort for a modified OMR was protracted and eventually terminated. Thus, no OMR is being used in the 1986 pretest, although testing data capture alternatives was a major objective of the test.

In November 1985, the Bureau decided to terminate its consideration of OMR technology as the primary data capture technology for the 1990 census. As a result of its decision, the Bureau ruled out the possibility of exploring the usefulness of this equipment, the new technology being considered for the 1990 census. Thus, the Bureau limited its options to the traditional FACT system and data keying.

From the beginning, the Bureau was aware that the commercially available OMR equipment did not satisfy all existing decennial census needs such as paper size and the use of a variety of marking instruments. Thus, testing the unmodified OMR equipment unless the Bureau was considering revising its requirements did not seem prudent. Absent this possibility, the Bureau should have formalized its requirements and initiated an effort to test a modified optical mark reader early in the decade.

CURRENT DECISIONS AND PLANS ON DATA CAPTURE

The Bureau's present plans on data capture and processing office configuration, as developed in late April, will result in a processing operation for most of the nation similar to that used in 1980. Our understanding is that the Bureau's FACT system will be used as the primary data capture technology. Filming of all questionnaires will be performed in 10 to 15 processing offices. The location of the film processing and scanning has not yet been decided. Questionnaires from hard-to-enumerate areas (possibly 12 percent of the nation) will be sent directly to these processing offices to be captured and where they may be automatically edited.

For the remainder of the nation, questionnaires will be returned to district offices, where they will be manually edited, similar to the 1980 census. Only after the questionnaires are "perfected" will they be sent in batches to the processing offices for data capture. In 1980, questionnaires were not sent to data capture until all questionnaires from a district were reviewed, "perfected" and batched, and the office was closed.

We have several reservations about the Bureau's late April decisions on data capture and the data processing office configurations. The Bureau's current plan for data processing is a hybrid system which incorporates concurrent data processing for a small portion of the nation and a modified 1980 data capture for the remainder of the country. On the one hand, we are pleased that the Bureau has decided on these basic automation activities 5 months earlier than originally planned as we have previously advocated, albeit without evaluation information from the 1986 pretest. We are also pleased that the Bureau has decided against using data keying, a relatively slow, error prone and expensive technology as a primary data capture technology. On the other hand, we are concerned with the decision to forego concurrent processing and revert to manual procedures for most of the nation because we believe the Bureau has thus foregone the benefits that could be derived from a more automated operation.

Sufficient details of the plans for the processing operations have not been developed to allow us to fully assess them. However, on the basis of the information obtained to date, we believe that the Bureau has compromised its goals for automation and for the census as a whole.

Because the majority of the questionnaires will not be captured until they are manually edited and "perfected," the important benefits of concurrent processing will not be obtained. Manual processes, particularly editing, will be used. Thus, some of the benefits from automation including speed and consistency and accuracy will not be obtained. Instead, a small army of temporary employees will probably be used. About 37,000 clerks were employed in the 1980 census to check returned questionnaires for complete and consistent entries. An early automated back up file will not be prepared. And premature destruction of the questionnaire forms as occurred in the 1980 census would remain a potential problem.

Maintaining controls over forms returned may be difficult if the questionnaires are sent to the processing office on a piecemeal basis. If information, such as population count, is not manually recorded from the returned questionnaires in the district offices, it also may be difficult to perform some coverage improvement programs, such as local review or recanvassing neighborhoods.

Additionally, we believe that the use of manual processes in most of the district offices and the use of 10-15 processing offices as compared to 3 in the 1980 census will help drive the per household cost, exclusive of inflation, of the 1990 census beyond the cost of the prior census.

We intend to monitor the continuing developments on data capture and processing office configuration because of the importance of the data processing operation in a decennial census.

This concludes my remarks, and I would be happy to respond to any questions.

RESPONSES TO FOLLOWUP QUESTIONS
MAY 15, 1986 HEARING BEFORE THE
SUBCOMMITTEE ON CENSUS AND POPULATION
COMMITTEE ON POST OFFICE AND CIVIL SERVICE

Q. 1. In your testimony, you recommend that the Bureau test a shorter form than has been used to date. Specifically you advocate dropping most of the housing questions from the 100 percent questionnaire. You based this recommendation on a finding that federal agencies don't use the 100 percent data. We have heard that many local governments do use this data for planning and in connection with federal grant applications.

-- Have you looked into the way local government use the 100 percent housing data?

-- What would you say to a Mayor or City Administrator who told you that they need the 100 percent housing data in order to file applications for federal discretionary grants?

A. Our work was generally directed to federal agencies because they are major users of census data and the Census Bureau had obtained justifications for its questions from them. In our work to date we have done very limited work on determining the use made by local governments of 100 percent housing data. However, on the basis of reviewing grant legislation and our limited work on local government's use of census data, we believe that data are rarely needed at the block level for grants. Sample data satisfies data requirements at higher geographical levels.

In responding to a local official who claimed to need 100 percent housing data, we would explore with the official his/her specific needs and try to determine the validity of the official's contention and whether alternative data might satisfy his/her requirements.

At the subcommitee's request we plan to do additional work to determine if there are valid needs for 100 percent housing data at the local level.

Q. 2. GAO strongly advocates that Census Bureau ought to carry out a test of the short form. What are GAO's expectations from such a test?

-- The return rates of the 1985 and the 1986 pretests were poor to fair. What return rates on the shortened simplified form would be acceptable standard for GAO?

A. GAO advocates the use of a streamlined short form to determine if it can achieve several advantages including a greater mail response rate, more complete questionnaires, a higher quality of data, possibly more cost efficient data processing options, and a reduced respondent burden. A greater mail response rate, less incomplete questionnaires and a more cost efficient data processing option, all should result in a less costly census. The Bureau's estimate of savings of $6 million for each 1-percent increase in mail response, basically because of reduced follow-up, is a good criteria in measuring the value of a streamlined form. Thus a 2-panel test, comparing the mail response from a regular short form to a streamlined form, would go a long way in determining the value of a streamlined form. In addition such a test could be used to compare the number of questionnaires that do not meet the Bureau's criteria of completeness. The Bureau follows up on incomplete questionnaires to collect the missing data.

Besides the cost of doing follow-up work, the interaction of a Bureau temporary employee injects a factor of enumerator bias into the data collected. This affects data quality. Moreover, removing the housing questions from the short form for the 1990 census would result in 85 million households answering eight fewer questions. This would reduce the respondent burden.

Q. 3. For the past several decades, the census planners have assumed that while there is a limit beyond which the census form could not go, generally up to that limit, the basic cost of getting to the household and getting the form back was not affected by the length of the form. Today you have testified that this assumption is wrong. Could you explain the evidence for your assertions?

-- What specifically leads you to believe that a very short form will greatly improve the response rate and reduce the cost of the census?

A. In the absence of a specific test, we do not know the extent to which a streamlined form will increase the mail response rate. However, we have evidence that indicates there would be an increase in the rate. We have examined the mail response rates from the censuses in 1970 and 1980 and pretests in 1985 and 1986. They have indicated that a higher response rate is obtained from the short form compared to the long form. For example, in the 1985 and 1986 pretests there was a 6 to 10 percentage points higher response rate to the short form compared to the long form. The difference in response is particularly evident in the hard-to-enumerate areas. We have also reviewed results from the focus group studies in the 1986 tests. Participants in those studies, mainly from lower income groups, have questioned the need for data aside from information needed for the count. As discussed above estimated costs can be attributed to follow-up work. In addition more cost efficient processing methods can be considered.

Q. 4. What are some data capture techniques the Census Bureau ought to look into even now to prepare for the 2000 Decennial?

A. We believe that data capture should be closely integrated with the data to be collected. The amount of data to be collected and the type of information such as numbers and names have a significant bearing on the data capture technology. Possible data capture techniques to be considered for the 2000 decennial include data imagery, a relatively new technology, hand-held personal computers, and direct telephone imput to computers. These latter two methods might help pave the way for a paperless census.

Q. 5. Please list some of your recommendations to enhance long term planning cycle for decennial censuses.

A. In order to improve the Census Bureau decennial planning cycle the following could help

--a long range or strategic planning group should become a reality.

--a strong research and experimental program be included in the current decennial census, and

--a permanent core group with sufficient authority be maintained in the Census Bureau to provide continuing direction for decennial planning activities.

Mr. Garcia. Thank you, Mr. Dodaro.

What I would like to do is to ask the Bureau of the Census their views. Just let me ask this one question first to the Census Bureau. We are going to hear from the National Computer Systems representative that they are prepared to process a form that had more than one page. What they are going to do was to slit the form, process it, and then reassemble it. This is what they already do for a number of OMR documents that they process. Why was it unacceptable to the Bureau of the Census?

We are going to hear from Mr. Gail Franke in just a little while, but I would like first to get a response from the Bureau.

Ms. Miskura. Mr. Chairman, controlling 100 million forms, of which 20 million are likely to be long forms, is a massive operation, even under the best of conditions. Requiring additional operations to slit, process, and reassemble the forms would add to the risks, in terms of timing, costs, and the accuracy of the data.

Small problems with missing forms or pages stuck to each other could cause us major problems. The recent processing decision that we made will utilize the cameras—the page-turning cameras—that we had, like in 1980, which performed very well for us, and which did not require us to take the booklets apart.

We understand that NCS has that capability, but we think we have minimized our problems overall and will be able to control the Census much better using the FOSDIC cameras.

Mr. Garcia. You have heard GAO testify. There are a number of areas in which they have some, I believe, constructive criticism.

I guess the first one is the question of the cost.

Do you have a problem with what they said?

Ms. Miskura. We have considered cost in all of our developmental work and our decisions about automation for the 1990 census. We will only use those automation techniques that we feel will be cost-effective or improve the accuracy and timeliness of the Census.

While it is very difficult to attach an estimate of savings to an individual automation decision, we do believe that the overall system that is designed will be most cost-effective.

We will be using a number of automation techniques that we certainly feel will be cost-effective. In particular, the automated geographic support system, or TIGER system; the automated address-control file; an automated check-in, which will be used throughout the country; automated edits in some parts of the country; and the earlier data capture, much earlier than in 1980, for the whole country also; as well as our automated management information system.

Mr. Garcia. Yes; as you know, I chaired the Census Subcommittee from 1979 to 1982, and I remember going back to the Appropriations Committee for additional funding to complete the 1980 census. We were assured that the costs of the 1980 census would be contained and we would not have to do that. But as it turns out, we had to do that.

You know what happened just recently with all the publicity on the escalating costs of the space shuttle. They used all sorts of formulas to say What they said because they are going to have more space shuttles taking off. They said that the large numbers will more than compensate for the cost. I just think that we are living in a world today, and its true especially here in Washington, where we need to call things the way they really are and not the way you want us to hear them.

It just seems to me that it is going to be very difficult, I believe, in 1990 to go back to the well.

Why don't we just start from the beginning and really say that there may be some overruns, and deal with it in that fashion?

Mr. BOUNPANE. Mr. Chairman, I would like to say a few words about that because I think that you have stated the problem correctly.

If you take a look at the census costs from 1950 on, and remove inflation from them to put them on a common base, the cost per unit has gone up, census by census by census.

That, I think, comes from two reasons, the first being that the census changes over time. We try to do a better job with each census.

The second reason is that it gets harder to take the census over time in our complex society. And so, yes—to then say that that line of increase is suddenly going to stop and level off between 1980 and 1990 is an extremely ambitious goal.

In trying to cost the 1990 census we tried to seek the appropriate balance between asking for funds in a very restricted budget environment and seeing what we could do about modernizing the census to make savings in some places to pay for increases that might exist somewhere else.

I think you are correct when you say that there are possibilities that we are not going to be able to meet that goal, and some of us have concerns about that as well.

What we want to do now is try to produce the cost of the census in a much more detailed manner, now that we have made some very basic decisions, to see what we think that total cost will be. If it is more than the stated goal we will come to the Congress and ask for those funds in advance. Because as you pointed out, the situation we experienced in 1980 was not a good one. We, like you, do not want to experience that, in terms of coming in for additional moneys at the 11th hour and all the associated problems with that.

One potential way to handle that is to have some kind of reserve set aside for unexpected difficulties. For example: Budget the census at an 80 percent mail return rate, but have a reserve set of funds available such that if, for some reason, the mail return rate in 1990 should come in at 75 percent, there are moneys available to attack that problem.

If I could add one more thing, though, about automating the census, and the cost of it. It is not necessarily true that automation saves money in a census environment. It is very true in an ongoing operation. If you can automate routine tasks, that should save money over time, because you have the chance to amortize huge investment costs of equipment.

Unfortunately, in a census the opportunities to do that are relatively limited. You have to pay a large amount of money to develop machinery and to purchase it. You use it only for a very short amount of time. And though it may help you save money by eliminating a labor-intensive task, on the other side of the scale is the huge amount of money you had to spend to develop and procure it. And those do not necessarily balance.

To me, the real advantage of automation comes in better control and better quality, which is important; not necessarily lesser money. Sometimes those facts get a little bit confused, and I thought it was worth putting that on the table.

Mr. **Garcia**. I see that Mr. Dodaro would like to respond to that. It is very important that we work together on this. It is good to have somebody looking over your shoulders. Constructive criticism is going to benefit us in the long run.

Mr. Dodaro.

Mr. **Dodaro**. Mr. Chairman, I think Mr. Bounpane makes a number of excellent points. But I would like to point out that the household cost between the 1970 and the 1980 census more than doubled, even after excluding the increase in inflation.

I think that it presents a number of challenges to the Bureau, and will not be achieved without taking some risks and challenging the status quo, and may be more than the Bureau has been willing to do in the past. This is somewhat understandable, however, since it is a large operation with a lot of components.

There is a long lead time for the 1990 census, and the key is to maximize the use of that time to carefully explore some of these options. The time has not been used properly in the past, to the best advantage. Some of the testimony that you heard on May 1, from the academic community, and some of the testimony that you are probably likely to hear from NCS is along the same lines.

While it has not been conclusively proven, as Mr. Bounpane points out, that automation will save money, it has been proven over time that reducing the number of people needed to take the census, particularly the training associated with bringing in thousands upon thousands of temporary employees, would be cost effective.

The direction that we need to move in is the one that you are pushing the Bureau in—is to really challenge the status quo and to be more aggressive in pursuing some of these options, to try to take hold of the situation and be more assertive.

Mr. **Garcia**. Just let me say to the panel from the Bureau of the Census that—I think we have to be a little imaginative. Yes, we do have the census that comes at the beginning of each decade. Why can't we, during the course of the other 6 or 7 years be imaginative and go out and market that system. And in-house, use that system for other agencies or whatever the case might be. Even going out into the private sector. It seems to me that the Bureau of the Census, which I have tremendous respect for, and which, I believe, most people who know your function have respect for, would be only too happy to work with you in a way in which you could be out marketing your product. The current system that you have there, which you are concerned about, would have a tremendous cost overrun.

We have to be a little imaginative. And I think it would be very productive. Go ahead Peter.

Mr. **Bounpane**. Is it OK to add something here——

Mr. **Garcia**. Yes.

Mr. **Bounpane**. I think Gene made a good point. Many of us wish we had, perhaps, done a little more testing in the 1980 census to help us toward 1990. At this point in time we can't change that, but we certainly can learn from that and say "test more in 1990 to learn about the future so we are not faced with this problem come 1993 or 1994." That is an excellent point, and I think we hear that clearly.

Also, I heard what you said, Mr. Chairman, about being a little more imaginative about things and I think we can take that back and see what opportunities we have for that, as well.

I would like to point out, however, that automating the census—we always talk about it with regard to the capture of the information from the questionnaire into the computer. That is only one part of the total census process. There are many others, and we have, I think, been imaginative in those areas. The program we have for automating the maps is a major advance over what we did in 1980 and is in direct response to a problem that existed in 1980.

What we have already accomplished in terms of automating the address control file, and allowing for automated check-in through bar codes, is again, a direct response to a problem that occurred in 1980, where that was done manually, and using computers to solve a problem that existed then.

The other point I would like to make is that for all of its negative appearance of being technology that's been around for a while, the FOSDIC system in 1980 worked very well. The field operations lasted some 2 to 3 months, in some cases 4 months, longer than anticipated in 1980. And yet, we still had to meet the date of producing the counts to the President, by December 31, the last day of the year.

The FOSDIC system made up that time difference. That is, we were able to read all 88 million questionnaires in 1980 through that system in several fewer months than we had originally planned. Now, I do not say that that makes it better than anything else that ever was, but it did operate well in a difficult environment. And that is a hard bit of experience to overlook. Notwithstanding the need to, perhaps, move forward where we can find better ways to do things.

So, I think we have moved on some things, particularly where there were problems in 1980 but, perhaps, to some people's point of view, not as far as we could have on some others.

Mr. **Garcia**. There is a vote on the floor, and it is an important vote, so what I would like to do is just call a quick recess. I'll run right over and I'll come right back.

[Recess taken.]

Mr. **Garcia**. Peter, I think there were some other parts of that slide presentation by GAO on which I saw you taking notes. Perhaps you would like to try to respond to some of the thoughts of the GAO.

Mr. **Bounpane**. Mr. Chairman, Susan will make some and then I will do some others, OK?

Mr. **Garcia**. Fine.

Ms. **Miskura**. I think we would like to respond particularly to the next-steps portion that Gene concluded with. Regarding the requirement requiring testing of a short short form, in effect, the feasibility of doing that really is dependent on the need for the data. Particularly, GAO has expressed concerns about the housing questions.

If we felt that they were not necessary on a 100-percent basis, we would take them off of the form, and it really would not be necessary to test that.

The reason those questions are included on the form is because in our continuing conversations with a broad variety of data users, we are being asked, and they are justifying the requirements for those data at small area levels.

We also feel that a number of the housing questions are necessary to assure coverage both of the population and of the housing inventory.

We realize that GAO's discussions with some Federal-level users have come up with information that is inconsistent with the information that we have collected. We are planning, on the basis of the information they obtained, to go back and talk to the members of the housing interagency working group at the Federal level, and also representatives of State and local governments to have them reassess their need for the 100-percent housing data.

Mr. **Garcia**. Just let me interject at this point, and it relates to my meeting yesterday with GAO. I have some problems with some aspects of the GAO's criticism because we have a special need today for the housing data the GAO proposes to eliminate. Since 1981 we have had a steady decrease of construction of housing for the poor and for the lower income people. Because of that there have been many instances in which, I know for a fact in my city, there have been families who have been doubling up.

Now, it is based on that that I have some problems with the GAO's criticism because we need that data. There is no question we need some of that data.

There may be some census data that are questionable, but in terms of the number of persons who are residing within a particular apartment, as in the case of public housing in New York, I think that the housing data is absolutely essential. And it just seems to me that we are going to have a very difficult time getting that information anyway. Many of the people are not going to willingly come forward to volunteer information that there is an uncle or an aunt, with children, living there. These people have no other place to go so they have to doubleup.

So it is based on that that I feel that the housing questions—some aspects of it—are extremely important.

I think counsel would like to ask a question at this point.

Ms. **Fernandez**. This question is directed to the Bureau. One of the concerns that we have, regarding the use of the short form, is particular to the housing questions. But it appears from the testimony, though you talk about the respondent burden, that it may not necessarily be so, as GAO suggests, that there be some respondent burden. It seems as though there is a tapering off of responses within a particular timeframe of the questionnaire.

For example, someone may start response but at the tail end, which are the housing questions, are not responded to. And that generates the additional cost of sending in an enumerator to get that information.

Have you looked to alternatives? For example, if the housing questions were removed from the short form, to increasing the survey, the 20-percent form additional housing questions where the quality of the data, the release of the data, would be earlier. Would that not offset some of the concerns that the data would be—that the questions be asked on the 100-percent count?

Have those other options been explored by the Bureau?

Mr. BOUNPANE. Yes. If I could answer that question plus something that Mr. Dodaro said earlier that sort of confused me—I think they are related.

Some of the questions, the housing questions, are needed for very direct reasons, like defining "is this a housing unit, or not"; and secondly, for some very basic planning that is necessary to know what a city should do, in terms of improving the situation and making changes.

We can argue about whether it is four questions that are really needed, or six questions that are really needed, or eight questions that are really needed. We can have that discussion, but it is about the number of questions, not the validity of the data need.

One thing that sort of surprised me, and maybe I misheard the GAO this morning, was a recommendation that all eight could be moved to the sample. We have never investigated that, moving every housing question to the sample questionnaire. Surely some are required on a 100-percent basis.

Some of them could be moved, I think. Not all eight, but I think we can take a good look at two or three of those to see whether or not they really are required of everyone, or if we could do just as you suggested, put them on the sample form, so that they would only be asked of 20 percent of the people.

I would only like to point out one thing about that, and that is looking at error rates from 1980, that is, the number of questions that were not answered that should have been answered. That error rate is worse on the long form than it is on the short form. That follows logic, of course. You would expect that. So that the more questions you put on the sample, the more likely you are going to get a nonresponse which could, in fact, increase field work anyway.

As Susan was pointing out, the key issue is the need for the information. If it is true need, and a true need in a small area like a block, then there is no alternative but asking it of everyone.

If the need is not for that small geographic area, but could satisfy for just the city as a whole, or a State, then surely asking it on a sample is more appropriate.

I think there are, perhaps, two or three of those housing questions that we should look at carefully again, to solve this bit of a dilemma that GAO's request to the Federal agency got a different answer than our request to the Federal agency. We have to solve that and say who is right. And if it is really not needed on a block area level then certainly moving it to the sample is something that should get serious, serious consideration.

Ms. FERNANDEZ. I know in previous hearings we talked about the questionnaire content and looked at the need for data. GAO in its testimony is raising questions not only regarding the need for the data on the 100-percent basis but also what data, in fact, the agencies are receiving. It appears that the agencies are getting survey information, sample information, and that appears to be adequate for the Federal needs. The GAO analysis is limited to a review of Federal users, excluding State and private data users.

Has the Census Bureau ever done an evaluation? I know at other hearings you responded with Federal legislative requirements of data users and other information like this—but the actual evaluation of where the need is and how that need correlates to the level of information—at the block level, the tract level, the county level—and done a thorough study on that?

Has the Bureau done that recently? Or at any point?

Mr. BOUNPANE. Not systematically, to my knowledge, at all.

Mr. DODARO. That is an excellent point that has been raised. There is a question of need but there is also a question of what is the most cost-effective way to meet that need; and what are some of the other options that could be available to satisfy that need rather than asking it of every household in the questionnaire.

I will ask Mr. Donoghue to elaborate a little bit further, but in terms of the points that Mr. Bounpane was confused on, I would like to eliminate some of that confusion. The questions that he is talking about, that are needed for coverage purposes and defining the housing unit, are ones we believe should remain on the form. It is the other questions that we think could be taken off.

We found there is a distinction between justification or desire for the information and what is actually used and the knowledge of the people from what they are actually using. Some people believed they were using 100 percent information, but in reality it was the public-use tapes from the long-form questionnaire.

So I think it really requires some probing and study to get behind some of the questions in terms of how they are using the information and what are other ways that those needs could be met without asking every household in the country to respond to how much rent they are paying, for example. These are the areas that we are suggesting that ought to be revisited and reexamined.

Ms. FERNANDEZ. One of the things that concerns us, and we discussed this previously in regard to the GAO proposal for the short form, is that moving the housing questions would appear to make it more difficult for data users to look at very specific characteristics of the communities, vis-a-vis their housing needs.

For example, if you are looking at the aged population, because the data is not in the 100-percent count, it may be difficult to determine what their specific housing needs are in particular States or counties.

Has the General Accounting Office explored what the loss of the data would be, or whether those characteristics across the Nation can be maintained by sample survey form?

Mr. DODARO. I would like to have Mr. Donoghue elaborate on this but I will give it a general stab first.

Many of those characteristics of information are embodied in a lot more detail on the long form questionnaire, which could be sampled. And I would like to, also, reiterate the point that it is a 20-percent sample which is used to make estimates at the county and track level, which are pretty good-sized levels and it is a good-sized sample of the country.

Mr. **Donoghue**. I think Ms. Miskura and Mr. Bounpane could probably respond better toward the loss of precision, but certainly at the State and county level the estimates from sample data would be sufficient to meet most State and local user needs.

What we are talking about here is how many people have to have the data at the block level, and have to use that data to make decisions.

The other aspect of our recommendations would not impact the size of the long form since all the questions on the short form are already on the long form, so if you took them off the short form the long form would still be the same size.

Ms. **Fernandez**. Just one other point. The General Accounting Office has proposed short-form testing. At previous hearings, the Bureau has stated that the New Jersey pretest, as applied, was a test of the short form. The GAO says that it was not.

Would the Census Bureau be amenable to testing a short form as suggested by GAO without the housing questions, either in a pretest or in some other mechanism to test whether or not, as they suggest, the response rate would increase?

Mr. **Bounpane**. I think we can certainly look at that and see if there is an opportunity to try to test that, keeping in mind that if the information is really not needed at small areas, then testing that would give you some information. But I do not know what you would do with it if you still had to produce that information for small areas.

It seems to me that the first question has got to be answered, and we would certainly support some kind of an investigation into what are the true uses of the information from the census and whether or not all of those housing questions really are needed for small geographic areas. We have not made that kind of a study in depth.

Mr. **Dodaro**. We would support that. I think Mr. Bounpane is right. You have to weigh the needs of the information against the benefits of it. But we would support doing both, so you have both pieces of information available.

In other words, pursuing the need question, keeping in mind options that could be considered to meet those needs at less cost. But also testing whether or not you do, in fact, get any benefits from increased mail response rates by using the shorter form. Then, having those two pieces of information, you will be able to weigh whether or not the benefits of increased mail response rates are worth the degree of precision and the need for some of the information.

Ms. **Fernandez**. It also may open up more automation options because the form would be shortened.

Mr. **Bounpane**. Let me just make one point about the sample size that both the chairman and you mentioned, because I think it is important.

Yes, 20 percent is a large sample by most standards, and it is easy to say that is a big enough sample size for information by tract or by county, but that is if you are talking about the entire universe.

Let me try to give you an example. Suppose you are interested in information about single, elderly women who rent their unit. Now, it is conceivable that a city would want information on how many women over 75 live alone and rent their unit. And they might want that information by neighborhood, not solely for the city as a whole. It seems to me a reasonable piece of information to want to have.

Now, at the present time you could get that when collected on a 100-percent basis. Alright? But if you asked renter and owner on a sample, rather than 100-percent, you can't get that information by very small areas. Alright? And you start to lose information very quickly for subsets of the population.

It is important to remember that, as well, when discussing the sample-size argument.

Mr. GARCIA. That is an excellent point.

I think what we have to do here is that there are many local governments that are going to need this information. There is no question about it. But I think that some of the criticisms that have been stated by the GAO deserve consideration by the Census Bureau. If you could put your heads together, I think we may come out with some sense of getting data that we could all live with, and from which we would get maximum use out of for the next decade, for the decade of the 1990's.

There are several other questions that have to be asked, but in fairness to our next witness Mr. Gail Franke, this panel will shop here. Before Mr. Franke comes on I would like to thank both the GAO and the Bureau of the Census for their participation.

Now, I would ask you to stay, because we may need you later on for questions and answers. Mr. Gail Franke, who is the vice president of Federal Government marketing, National Computer Systems, is going to testify now. There may be something which we will ask all of you when he finishes testifying.

STATEMENT OF GAIL FRANKE, VICE PRESIDENT, FEDERAL GOVERNMENT MARKETING, NATIONAL COMPUTER SYSTEMS

Mr. FRANKE. Mr. Chairman, on my left is Mr. Robert Roelf. He is NCS technical project manager for census-related activities. And on my right is Mr. Jeffrey Goldberg, NCS program manager for the Bureau of the Census.

I must say I was relieved to see you excuse the previous witnesses, Mr. Chairman. Between GAO and the Bureau is not exactly the place I wanted to be and that appeared to be the only open chair, so. [Laughter.]

This is much more comfortable, with all due respect to my good friends from both those agencies.

On behalf of myself and National Computer Systems we appreciate the opportunity to appear before your subcommittee today.

I would like to offer some summary comments on the material previously submitted to the subcommittee and these comments will address four issues.

First, a brief, but I believe pertinent, statement describing NCS, the corporation, which is to fundamentally establish our credentials in the optical mark reading arena.

Second, an abbreviated description of our involvement in the OMR evaluation process conducted by the Bureau, and our assessment of the results that obtained therefrom.

Third, our observation of the Bureau's overall planning process.

And last, some general recommendations relative to a research and development program which we feel the Bureau should implement immediately.

To get to the first of those issues, since 1962 National Computer Systems, through its OMR systems, has been providing technological solutions for the large scale data collection needs of education, government, and industry. As a corporation, NCS in the fiscal year just ended generated revenues in excess of $215 million, and maintains a work force of nearly 2,000 employees nationwide.

While OMR systems were first pioneered in the area of educational measurement, they have since been successfully used for many other large scale data collection tasks including surveys and assessments, application processing, health care reporting, and order entry, to name just a few.

In addition to the manufacture of OMR systems, NCS is also one of the largest processors of OMR scannable documents in the world. Receiving annually over 130 million scannable forms at its Iowa City, IA, facility, NCS has established industry standards for the logistics management of large volume data collection projects.

NCS's experience as a supplier of systems and services to the Federal Government has also been extensive. For instance, we currently process approximately 7 million applications for student financial aid through our information services division, and we provide distributed, as well as centralized, OMR data entry systems to many agencies of the Federal Government. NCS currently has over 1,000 such OMR systems installed in various Federal installations around the world.

Finally, we have transported this technology to foreign markets. In the international sector, NCS has again served the complex data collection needs of a variety of areas.

It was in the international arena that NCS gained its initial experience in supporting census data collection. NCS OMR systems have been used to conduct the population censuses of Venezuela, Chile, and the Dominican Republic, and the Mexican agricultural census. The experience gained in supporting international census data collection has served to educate NCS to the special requirements of census processing.

It was during the period of time when NCS was supporting foreign census programs, that we engaged in initial discussions with the U.S. Census Bureau, regarding the possible role OMR could play in processing a U.S. decennial census. Between 1980 and 1984 several delegations of Census officials traveled to NCS facilities to view NCS OMR systems and talk to NCS engineers regarding the application of that technology in the 1990 census.

During this same period, NCS personnel toured various Census facilities in order to gain a more detailed understanding of U.S. Census requirements for 1990, and to understand more fully the Bureau's resident data capture technology, FOSDIC.

In 1984, the Bureau acquired an NCS OMR system, the W-201 for testing and evaluation. In making the commitment to test this system, we believe the Bureau recognized two of the major benefits that we feel can be derived from OMR processing. The economy and efficiency of direct source document data capture, and the transportability of this technology.

NCS worked very closely with the Bureau on this evaluation program, focusing on the Tampa pretest to be conducted in March 1985. Two significant issues surfaced very early on in this activity.

First, there was the matter of the form. As indicated earlier, NCS proposed a four-page booklet for the census pretest form, which would be slit for scanning, and if desired, automatically reassembled after scanning; a methodology used by several major OMR processors.

The Bureau disagreed with this proposal and directed us to accommodate all short-form data on a single 8½-by-11 sheet. The result was an admittedly dense document which the Bureau and NCS agreed would have to be larger and more open in future use.

The second major issue was the No. 2 pencil, the industry standard OMR marking instrument. While it was marginally practical to insert and mail pencils for the Tampa pretest, some 120,000-odd households, it was clear that a much wider range of marking instrument would have to be scannable in order for OMR to be used as the primary data capture technology in 1990.

In January 1985, and as a direct result of preparing for the Tampa pretest, we submitted an unsolicited proposal to enhance standard NCS technology to produce an OMR system which we believed, and were prepared to demonstrate, would meet the census-specific requirement.

This proposal addressed the need for a larger form, the use of a wide range of marking instruments, and several other significant enhancements. This enhanced system was to have been delivered in two phases. The initial system, including the larger form capability, in January 1986, for use in the Los Angeles pretest; and the remaining enhanced features in September 1986, consistent with the Bureau's then operative schedule for technological evaluation.

The Bureau, after evaluating NCS's proposal, and reviewing the results of the Tampa pretest, concluded that it was beneficial to continue the OMR evaluation process. Accordingly, in June 1985 they announced their intention to contract with NCS for the acquisition of an enhanced system.

However, after having received some level of response from some other interested parties, the Bureau reversed their decision and announced that they instead intended to conduct a competitive procurement.

In September 1985, the Bureau released a draft specification for comment. NCS submitted comments, having been assured of a prompt release of the RFP. However, in early November 1985 NCS was informed that no RFP would be forthcoming, and furthermore, that all OMR evaluation, as related to the 1990 census had, in fact, been discontinued.

The Bureau gave three reasons for this action.

First, on the basis of the Bureau's assumptions, the Bureau's own FACT 90 system had a more favorable cost-benefit ratio than did the OMR solution.

Second, the Bureau believed it had run out of time, because in order for the Bureau to manufacture sufficient FACT 90's they could not wait until the fall of 1986 for the decision, as per the original schedule.

And third, the Bureau took the position that the enhanced OMR system they would have tested in 1986 would not necessarily have convinced them that OMR could support the 1990 census.

NCS, of course, took exception to the Bureau's stated reasons for suspending the program. We still continue to believe that the direct data capture capability, coupled with the ability for wider deployment, serves to more cost effectively support the dual roles of concurrent processing and accelerated turnaround. OMR is a much simpler process where in most instances should equate to less cost.

With regard to the time constraints, NCS felt that there was then ample time to conclude a substantive and complete evaluation of OMR and to meet the production requirements for 1990, in that the system NCS proposed to deliver in 1986 would have given the Bureau the facility to thoroughly test the entire enhanced OMR system.

In stepping back from the specifics of the Bureau's OMR Technology Evaluation Program, there are several observations that we feel we can make.

The Bureau has a long and proud history as a pioneer in the use of advanced technology and innovative methods. Recently, however, the Bureau, it appears, has had more difficulty making significant technological advancements.

In planning and implementing a decennial census, the Bureau is continually having to make decisions along a risk versus opportunity continuum. While the stakes are high in conducting a decennial census, it is our observation that the Bureau primarily tries to minimize risk, at the expense of seizing the opportunity to create fundamental, technological improvements in decennial census operations.

Part of the problem can be traced to the Bureau's planning process. It appears that the Bureau waits too long between decennial censuses to begin the process of identifying, evaluating, and inserting new technology into its operations.

Consequently, as census day approaches, the Bureau has tended to fall back on procedures and systems used in previous censuses because sufficient time no longer remains to successfully test, procure and implement innovative new automation systems.

The Census Bureau seems to be repeatedly caught in a continuous cycle of developing plans to move forward with its decennial automation only to be forced to fall back on existing data processing systems.

This is a problem that has historical perspective in that the Bureau has consistently tended to rely heavily on the use of its in-house capabilities for the development of any new or innovative methods or systems.

Early in the Bureau's history this was done out of necessity, due to the lack of commercially available alternatives. Today, however, the commercial sector provides a broad range of products and services which should be of direct benefit to the Census Bureau.

Finally, where should the Bureau be going for the 2000 decennial census? It is very clear to NCS, as it has been to many others who have testified before this subcommittee, that planning for the year 2000 should have begun yesterday.

The Bureau should implement a program immediately to develop a qualitatively superior data capture system. Such a system must combine the best features of currently available technologies, including FOSDIC, OMR, OCR, image capture, optical disc technology, online microfilming, and multipage booklet processing.

Additionally, the design philosophy of such a program must be flexible enough to accommodate applicable emerging technologies. The goal of this R&D program should be to develop a Census data capture system which would exceed all that is currently available when viewed from the perspective of cost effectiveness.

We believe very strongly that this project must be a collaborative effort wherein the best resources of the Bureau work closely with experts drawn from external organizations.

Additionally, the Bureau should continue examining and analyzing all emerging technologies, with specific attention directed to those technologies which could ultimately provide a truly electronic data capture capability.

Finally, we wish to reemphasize the importance of beginning this project immediately, with the goal of having a prototype system, or systems, ready for live tests during the 1990 census.

For this goal to be realized, however, the Bureau must make the R&D program of the highest priority. It is our opinion that this committee, through hearings such as this, can play a positive and supportive role in assisting the Bureau in achieving these goals.

That concludes my comments, Mr. Chairman, and we, too, stand ready to answer any questions you might have.

[The statement of Mr. Franke and his response to written questions follow:]

Testimony of Mr. Gail A. Franke
Vice President of Federal Government
Operations, National Computer Systems
Presented to the House Post
Office and Civil Service Committee,
Subcommittee on Census and Population

May 15, 1986

Mr. Chairman, on behalf of myself and National Computer Systems (NCS), we appreciate the opportunity to appear before your subcommittee today. I am accompanied by Mr. Jeffrey Goldberg, NCS's Program Manager for the Bureau of the Census, and Mr. Robert Roelf, NCS's Project Manager for Census-related activities.

Since 1962, National Computer Systems has been providing technological solutions for the large scale data collection needs of education, government and industry. As a corporation, National Computer Systems, in the fiscal year just ended, generated revenues in excess of $215 million, and maintained a workforce of nearly 2,000 employees nationwide. National Computer Systems is the world's largest supplier of Optical Mark Reading Systems.

An Optical Mark Reader (OMR) directly converts marks made on preprinted paper forms into machine readable computerized data. Over the last 24 years, NCS has supplied its OMR systems to a broad range of users. The earliest applications of Optical Mark Reading occurred in the field of educational measurement. Optical scanners have been used in this area for over 20 years to provide high speed scoring of standardized educational tests. As the educational and career options of millions of students have depended on the results of standardized tests, it has been critical from the outset that OMR systems provide high speed, high volume test scoring and, at the same time, maintain the highest standards of accuracy. While OMR systems were first pioneered in the area of education, they have since been successfully used for many other large scale data collection tasks including survey research, application processing, health care reporting and order entry, to name just a few. In addition to the manufacturer of OMR systems, NCS is one of the largest processors of OMR scannable documents in the world. Receiving annually over 130,000,000 scannable forms at its Iowa City, Iowa facility, NCS has established industry standards for the logistics management of large volume data collection projects.

NCS's experience as a supplier of systems and services to the federal government has also been extensive. We process about 7 million applications for student financial aid through our Information Services Division, and we provide large scale distributed OMR data entry systems to many agencies of the federal government, including the Air Force, Postal Service, FAA, OPM, and the Department of Education for "in-house" scanning of OMR forms. NCS currently has over 1,000 systems installed in various federal installations.

Finally, NCS has transported its OMR technology to foreign markets. In the international sector, NCS has again served the complex data collection needs of education, government and industry. It was in the international arena that NCS experienced its first success in supporting census data collection. NCS OMR systems have been used to conduct the population censuses of Venezuela, Chile, and the Dominican Republic, and the Mexican agricultural census. The experience gained in supporting international census data collection has served to both educate NCS to the special requirements of census processing as well as to clearly identify the major benefits which can be derived from using OMR systems for Census data capture.

During the period of time when NCS was supporting foreign census programs, it also engaged in discussions with the U.S. Census Bureau regarding the possible role OMR could play in processing a U.S. decennial census. Between 1980 and 1984 several delegations of census officials travelled to NCS facilities to view NCS OMR systems and talk to NCS engineers regarding the application of OMR technology in the 1990 Census. During this same period, NCS personnel toured various Census facilities in order to gain a more detailed understanding of U.S. Census requirements for 1990 and to understand more fully the Bureau's resident data capture technology, FOSDIC.

In 1984, the Bureau acquired an NCS OMR system, the W-201 for testing and evaluation. This model had been used successfully in several international census processing projects.

In making the commitment to test OMR, we believe the Bureau recognized two major benefits which would be derived from OMR processing. First, OMR systems provide direct source document data capture. This means that information is extracted directly from the paper form as opposed to indirectly, which occurs under FOSDIC's filming, developing and film scanning process. Using a direct source document data capture system, there is reduced clerical handling of forms which translates into a reduction in processing turnaround time and therefore, a reduction in cost. The second advantage of OMR is that the technology can be deployed, if so desired, in a highly decentralized fashion. In as much as the Bureau was interested in converting from the centralized batch processing mode which it employed in the 1980 Census to a "flow based" or "real time" processing scenario for the 1990 Census, it was critical that the data capture technology be able to be distributed close to subsets of the respondent population. By processing in relatively close proximity to respondent households, the Bureau would be able to accomplish its goal of concurrent processing and thereby more efficiently identify non-respondents and correct forms which failed various content edits. The end result would be more complete coverage and cleaner data.

After preliminary testing of the NCS scanner, the Bureau decided to use the OMR system to process the "short" form in the 1985 Tampa, Florida, pretest. Between August 1984 and February 1985, NCS worked with the Census Bureau to prepare for the Tampa OMR test. NCS assistance included advice and guidance relative to software development, field engineering adjustments to the scanning system, and support in forms design and printing.

With regard to forms design, the Bureau originally asked NCS (using the model of the 1980 Census short form) to do a design of a scannable short form for the 1985 pretest. In responding to the Bureau's request, NCS initially designed the form as a 4 page booklet. In order to process the document on the NCS scanner, the 4 page booklet would be slit into two 8.5"x11" sheets for scanning. This is a standard booklet processing procedure used by all major processors of OMR scannable documents. As an example, a major commercial user of NCS systems processes over 11,000,000 24-page booklets per year using this methodology. The Bureau stated, however, that this process was not consistent with their standard procedures. NCS originally proposed the 4 page document because it believed that, given the number of response items, it would be easier for respondents to complete a form that spread data items across four pages. In deciding not to slit the booklet, the Bureau required that all data items be compressed to a single two-sided 8.5"x11" sheet. While this was accomplished for the 1985 pretest, the actual document was rather dense when viewed from the respondents perspective. NCS and the Bureau agreed that the OMR form, if it was to be constrained to a single sheet, would have to be larger.

Another issue which needed to be addressed in the 1985 pretest was the use of No. 2 pencils to complete the Census forms. Commercially available OMR scanners have been designed to read marks made by No. 2 pencils. The historical basis for this requirement can be found in the fact that OMR systems were first used in educational testing environments where students were presumed to have pencils or could be provided with pencils at a test administration site. Other users of OMR systems beyond the educational sector never saw the use of a pencil as a limiting factor. As a consequence, it has never been a design requirement to equip NCS scanners with the ability to read marking instruments other than pencils. It is interesting to note, that where OMR has been used in international census taking, the actual censuses were 100% enumerator administered. Within such a controlled environment it was practical to supply enumerators with pencils. For the 1985 Tampa pretest, the Census Bureau decided to provide pencils to the 125,000+ respondents as part of the mailing packet. It was clear, however, that if OMR were to be used as the primary data capture technology for the 1990 Census it would not be realistic to provide pencils to 100,000,000 households. Therefore, the scanner would have to be enhanced so that it would read other types of marking instruments.

As NCS continued to work with the Bureau in preparation for the 1985 pretest, we began to see that, for OMR to provide optimum performance and benefit within the U.S. Census environment, certain enhancements/modifications would need to be made to the NCS standard commercially available product. Accordingly, in January 1985, NCS submitted an unsolicited proposal to the Census Bureau to provide an OMR system which would meet the specific and unique requirements of Census processing. In this proposal we presented a comprehensive plan for engineering the modifications which would be required in order to meet the Census Bureau's needs. Major tasks included the modification of the scanner transport in order to read a larger form (11"x17" as opposed to 8.5"x11"); enhancement of the optics system in order to reliably read marking instruments other than a No. 2 pencil (e.g., ball point pen, flair tip marker, etc.); and the development of an internal humidity compensation and skew detection and correction capability which would ensure the accurate reading of forms regardless of changes in the forms' dimensions resulting from variable environmental conditions.

NCS proposed to deliver this enhanced scanning system in two phases. The initial enhanced model would have been delivered in January 1986 in time for use in the 1986 Los Angeles County pretest. This system would have had the capability to read the large form. In September 1986, NCS would have delivered all other system features including variable marking instruments, humidity compensation and skew detection. At that time, the Bureau would have been able to complete its testing and evaluation of OMR in sufficient time to meet its then operative deadline for making the decision on the primary technology to be used for data capture in the 1990 census.

8

Between January 1985 and May 1985 the Bureau considered the NCS proposal. Several meetings occurred between NCS and Bureau personnel. Also, during this period (March 1985 - June 1985), the Bureau was evaluating the performance of the commercially available NCS scanner which was being used to process the 1985 Tampa pretest. From the NCS perspective it was essential that the OMR system employed in the Tampa pretest perform successfully in order for the

Bureau to be willing to undertake further testing of an enhanced OMR system. During the Tampa pretest, the OMR system performed very well, considering that it was a "standard off-the-shelf" product operating in a highly specialized environment. In essence, the pretest proved that OMR as a generic technology could successfully support decennial data capture processing. The pretest also confirmed, to no one's surprise, that the standard commercially available product would need to be modified in order to meet Census requirements. Indeed, the OMR evaluation report prepared by the Bureau's Technial Services Division validated the need for many, if not most, of the enhancements NCS had proposed to make to its standard product.

The Bureau's key decision makers agreed that the results of the OMR test in Tampa justified the need for further testing of an enhanced OMR system. Accordingly, on June 21, 1985, the Bureau announced in the Commerce Business Daily its intention to contract with NCS for the delivery of an enhanced OMR system to be used in the 1986 Los Angeles County pretest. In this CBD announcement, the Bureau requested that other concerns having the capability to provide the required OMR system respond within 30 days. In August 1985, NCS was informed that the Bureau had received responses from other interested parties and that, on the basis of these responses, it would not proceed with an award of a contract to NCS. The Bureau stated, however, that it would be issuing a Request for Proposal for the desired OMR system. In September 1985, a draft specification was released by the Bureau for comment. NCS submitted written comments to the Bureau in anticipation of the prompt release of the formal RFP. In early November 1985, NCS was informed by the Bureau that it had decided to cancel the OMR procurement and to discontinue any further testing or evaluation of OMR for use in the 1990 Census.

The Bureau provided 3 principal reasons for cancelling the OMR evaluation program. First, the Bureau believed that when comparing OMR and the Bureau's FACT 90 System from a cost/benefit perspective, OMR did not appear, on the basis of the Bureau's assumptions, to provide either sufficient cost savings or funtional advantage to justify further testing. Second, the Bureau believed that it had run out of time for making its technology decisions for the 1990 Census. Thus, the original schedule which called for a technology decision to be made in the fall of 1986 was too late. In order for sufficient FACT 90 systems to be ready for the 1990 Census, a decision relative to its use in the 1990 Census could not be postponed until the fall of 1986. Given that the Bureau felt OMR still required additional testing to confirm its viability for decennial census processing, and given further that the FACT system had been used for decennial data capture since 1960, the Bureau claimed that there would be no advantage to be gained from continuing the OMR testing program. Lastly, the Bureau believed that the enhanced OMR system which they would have received for testing in 1986, would not necessarily have proven the ultimate utility of OMR for decennial Census production processing.

NCS has taken exception to the reasons given for cancellation of the technology evaluation program. With regard to a cost/benefit comparison of

OMR and FOSDIC, NCS continues to believe that the direct source document data capture provided by OMR and the ability to distribute OMR technology widely in the field, serves to more effectively promote the Bureau's goals of concurrent processing and accelerated document turnaround than does FOSDIC. To be sure, OMR does not currently provide certain features available through FOSDIC. OMR processes multi page booklets (e.g., long form) by slitting the booklet, scanning the sheets and, if desired, automatically reassemble the booklet after the scanning process. Since FOSDIC forms must first be microfilmed before scanning, FOSDIC produces an archieval record of the Census form as a by-product of that microfilming process. Whereas under OMR processing either a microfilm capability would have to be added to the OMR system or microfilming would have to be done as a separate and possibly later process. On the other hand, OMR processes paper directly, whereas FOSDIC data capture requires a sequential process of microfilming, microfilm developing, and microfilm scanning. If a Census form fails edit under the OMR processing scenario, the correction can be made directly to the form and the paper directly rescanned. Under FOSDIC, however, if a document fails edit, after the correction is made on the form, the entire filming, developing and scanning process must be repeated.

With regard to the Bureau's need to accelerate its technology decision timetable, NCS can only state that it had been operating for over a year and a half under a schedule that assumed a technology decision would be made in the forth quarter of calendar year 1986. NCS felt that this schedule was reasonable when viewed from a development, testing and procurement perspective as well as gearing-up for production and deployment. The sudden change in schedule was very unexpected given the precision with which the Bureau had originally laid out its timetable and master plan.

Finally, the system that NCS would have provided for testing in the 1986 Los Angeles pretest would have allowed the Bureau to evaluate all manditory features and components. The system would have been a pre-production model that would have permitted the Bureau to determine whether OMR could operate as the primary data capture technology for the 1990 Census.

Not withstanding the Bureau's decision to terminate the OMR technology evaluation program, NCS believes the Bureau conducted this program consistent with the highest standards of professionalism and with a genuine commitment to a thorough and fair evaluation.

In stepping back, however, from the specifics of the Bureau's OMR technology evaluation program, there are several general observations that can be made. The Bureau has had a long and successful history as a pioneer in the use of advanced technology and innovative methods. Recently, however, the Bureau has had more difficulty making significant technological advancements.

In planning and implementing a decennial Census, the Bureau is continually having to make decisions along a risk vs opportunity continuum. While the stakes are high in conducting a decennial Census, it has been our observation that the Bureau tends to place its primary emphasis on risk minimization as opposed to taking advantage of the opportunities to implement major advancements in the use of technology for the decennial census. We believe that part of the problem can be traced to the Bureau's planning process. The Bureau seems to wait too long between decennial censuses before it begins to define its approach to technology evaluation. Consequently, as Census Day approaches, the Bureau has been forced to fall back on procedures and systems used in previous censuses because sufficient time no longer remains to sucessfully test, procure and implement innovative new automation systems. The Census Bureau seems to repeatedly put itself in a continuous cycle of developing plans to move forward with its decennial automation only to be forced to fall back on extant data processing systems, with some measure of Bureau developed enhancements.

This is a problem that has historical perspective in that the Bureau has consistently tended to rely heavily on the use of its in-house capabilities for the development of any new or innovative methods or systems. Early in the Bureau's history, this approach was necessary, based on the lack of commercially available alternatives. Today, however, the commercial sector provides a broad range of products and services which could be of direct benefit to the Census Bureau.

Where should the Bureau be going for the 2000 Decennial Census? It is very clear to NCS that planning for the year 2000 should begin today. The Bureau should undertake an immediate research and development effort. This R&D program should concentrate on developing a qualitatively superior data capture system. Such a system must combine the best features of currently available technologies, including FOSDIC, OMR, OCR, Image capture, optical disc technology, on-line microfilming and multi-page booklet processing. Additionally, the design philosophy of such a program must be flexible enough to accommodate applicable emerging technologies. The goal of this R&D program should be to develop a Census data capture system which would exceed all that is currently available when viewed from the perspective of cost effectiveness, efficiency and overall function. We believe very strongly that this project must be a collaborative effort wherein the best resources of the Bureau would work closely with experts drawn from external organizations.

Additionally, the Bureau should continue examining and analyzing all emerging technologies, with specific attention directed to those technologies which could ultimately provide a truly electronic data capture capability.

Finally, we wish to reemphasize the importance of beginning this project immediately. If the Census Bureau begins a research and development program this year, it is possible to have a prototype system ready for live testing during the 1990 Census. For this goal to be realized, however, the Bureau must make the R&D program a high priority.

National Computer Systems responses to the questions submitted by the Committee on Post Office and Civil Service, Subcommittee on Census and Population.

Question No. 1: As you know, the Census Bureau provides a back-up system to the birth and death records of the country. If they are ever lost or damaged, people can obtain vital documentary evidence by consulting a microfilm copy of their old census records. If the Census Bureau decided to switch to the OMR technology, how would they continue to serve this need?

NCS Response: During the period of time that NCS was working with the Bureau on the OMR technology evaluation, NCS had several discussions with Bureau personnel relative to Census archival record requirements. NCS understands fully the Bureau's need for a microfilm copy of the original census form. NCS has defined three possible alternative approaches to satisfying this requirement. Probably the most efficient and cost-effective approach would be to equip the OMR scanner with an in-line microfilming capability. NCS engineers have assessed the feasibility of mounting a microfilm camera within the OMR transport housing. Based on this assessment, we believe it is totally within the realm of possibility to equip an OMR system with a microfilming capability. It is assumed that document microfilming would occur as a continuous process following OMR data capture. However, instead of developing the microfilm within the processing office, since the microfilm copy is not essential to the data capture process itself, it is assumed that the microfilm cartridge would be shipped to the Jeffersonville, IN, Data Preparation facility for subsequent developing. The second alternative would be to have a separate microfilm camera at the processing office in order to create the archival record of the Census form. Under this scenario, after source documents were scanned on the OMR system, they would be passed on to a microfilming operation. Again, the raw film would be shipped to Jeffersonville for developing. Finally, the last alternative would be to ship the Census forms themselves, to Jeffersonville for microfilming after OMR data capture was completed in the processing office. We consider this last alternative least desirable based on the cost entailed in shipping all Census forms to Jeffersonville.

Question No. 2: In your statement, you say that each year NCS processes 130 million OMR documents. Just to place the Census in perspective, the decennial involves more than 106 million housing units, each of which will have a form. Furthermore, 85% of these forms must be processed in less than 10 weeks -- a very short period of time. It looks like the job of processing the Census with OMR could easily require equipment that involved several times the capacity of your firm. I have two questions about this:

First, would it be possible to obtain this amount of OMR equipment?

Second, what would the government do with the equipment when we are finished with it? Could it be leased so that we would not have to absorb the full cost of buying it?

NCS Response: A. Yes. NCS's manufacturing capabilities are more than adequate to meet the OMR scanning needs of the Decennial Census. Had the Bureau stuck to its original timetable for technology evaluation and procurement, the requisite number of scanning systems would have been ready for the 1988 dress rehearsal and the 1990 Census.

B. NCS has had several discussions with the Census Bureau regarding alternative procurement approaches for the required quantity of OMR scanning systems, and was prepared to address that issue in contract negotiations. We have always assumed that a combination of a lease and purchase arrangement would be the approach best suited to the Census Bureau. Under this approach, the Census Bureau would purchase a certain number of OMR systems to be retained as residual technology after the 1990 Census. The remaining systems required for decennial Census processing would be leased to the Bureau for the duration of decennial data capture. NCS believes that there would be an international market for the leased systems among countries conducting their censuses between 1991 and 2000.

Question No. 3: The Census Bureau mentioned in their testimony that one of the reasons for not going for the OMR technology is the issue of cost related to adopting the OMR technology to the decennial processing. How much do you estimate the cost would be or if you cannot estimate the cost, what are the cost factors to be considered?

NCS Response: In order to provide a definite cost estimate, we would need a set of comprehensive and stable requirements from the Census Bureau. One of our major frustrations has been our inability to acquire a firm fix on what the Bureau's requirements would be. However, the principal cost factors for which we would need a statement of Bureau requirements are as follows:

- Number of processing centers
- Number of 100% and sample forms to be processed per center.
- If the Bureau would agree to document slitting and, therefore, OMR scanning of the sample form, what is the page length of the sample form?
- What level of redundancy would the Bureau require at each processing center?

- What is the timeframe in which the Bureau wishes to process 100% and sample forms?

- How many labor shifts would be utilized at each processing center?

- How many machines would the Bureau intend to purchase, and how many would they wish to lease?

- What type of maintenance program would the Bureau wish to contract for?

0033G

Mr. **Garcia**. Before we do that, what I would like to do is ask both Mr. Dodaro, Ms Miskura, and Mr. Bounpane to come on back.

I would like the Bureau of the Census to respond, especially to the last statement made by Mr. Franke about the R&D, and how do we go about making that decision fast enough and quick enough for the 1990 census.

Ms. **Miskura**. We certainly have learned, from our attempts to plan the automation of the 1990 census, the amount of leadtime that it takes to explore the technologies that are available, to understand the possible technologies that will be available between the time our planning starts and when we actually have to take the census.

Those lessons certainly have told us that we do need to start research and development of data capture technology and other automation features of the year 2000 census as soon as possible.

It will be a Bureau priority to start defining the year 2000 census and automation requirements as soon as possible. We would request funding for that and we will use the opportunity of the 1990 census for research and experimentation with possible methodologies and technologies.

We have long thought about the ideal Census being a paperless Census based on electronic means of having data available on the computer and we do intend to work towards that end.

Mr. **Garcia**. Mr. Dodaro.

Mr. **Dodaro**. We certainly would support moving out in the direction that has been suggested. And in terms of planning for the year 2000, we have continually advocated the need to make maximum use of leadtime and think that is an opportunity.

The only caveat I would add to that is that it still may not be too late to include certain things in the 1990 census. And I would not want to see us move out with the idea of writing off the 1990 census and moving just totally, exclusively into planning for 2000.

I think the timing that Mr. Franke suggested about using the 1990 census as a test, is a good idea.

Mr. **Garcia**. Why can't we do that? If we can't do it nationally, why can't we do it regionally? Why can't we just try?

Ms. **Miskura**. To the extent that we identify potentially useful technologies and methodologies we will have the opportunity in the 1990 census to try them, perhaps, for—within a processing office or within a set of collection offices.

To the extent that can learn about the technologies and to the extent that we feel that it is appropriate to try to learn about them in the context of the decennial census our plans are to do that.

Mr. **Garcia**. I was flown out to Iowa City to look at the facility in 1979, just prior to the 1980 census. Now 1986—and we are still talking.

I have all the respect in the world for the agency, but—I guess some decisions are not being made when they should really be made. At least we should try—so we try and we lose—we will try something else; the idea is to really try and be ahead.

I don't think you should be just on a par. I think the Bureau of the Census should be so far out in front that everybody is looking to them for the leadership. I mean that. I say that with all due respect to the agency.

Again, the Census Bureau—stated in their testimony why they chose not to use OMR technology for the 1990 decennial. One of their reasons is that the scanners are sensitive to temperature and humidity. Another reason is that they experienced difficulties in evaluating the scanner accuracy.

Any comments?

Ms. **Miskura**. In our defining the parameters of our interest in the optical mark recognition technology early in our planning process, and our funding for planning the 1990 census officially began in fiscal year 1983, what we had hoped to do was to apply a principle that I think has been talked about this morning, which is to bring data capture as close as possible to the source of data that we could.

The idea, therefore, was to try to develop—or see whether OMR could be developed to decentralize as far as the collection office, 400 to 500 different locations. We had to make sure that the machine was fairly robust, so to speak, because in identifying 400 to 500 different locations for collection offices across the country we might not be able to get ideal space all of the time.

Our looking at OMR in Tampa was really the first step in what we envisioned as an ongoing process. The first thing we wanted to know was would we be able to function with a minimum of environmental controls. We did feel that we got the answer to that question, at least in terms of the existing technology, that we would not be able to do that in 400 to 500 different places.

Mr. **Franke**. Mr. Chairman, may I respond to that?

Mr. **Garcia**. Yes, please.

Mr. **Franke**. I was not at the Jeffersonville pretest—my two associates were, and I would like them to comment in a bit more detail.

But I would submit that the problems experienced were not necessarily problems associated with the scanning system, but rather problems associated with the form. As with a lot of data processing media, including magnetic tapes and discs that are transported any distance, that media must be acclimatized before it is inserted into a data processing system. I believe that during the Jeffersonville pretest we were transporting documents from Tampa up to Indiana in March. So the documents underwent a rather severe temperature change and humidity change.

We suggested at acclimatization, I believe, in fact, the Bureau did acclimatize, but I would suggest that it may have moved the form than the hardware itself. And I think that too affected, perhaps, the accuracy of the system.

If either one of you would like to comment on that.

Mr. **Goldberg**. Mr. Chairman, the only additional comment I would like to make——

Mr. **Garcia**. For the record—your name?

Mr. **Goldberg**. Mr. Goldberg. Jeffrey Goldberg.

The only additional comment I would like to make with regard to the environmental issues was that it is true, we did recommend to the Bureau in the 1985 pretest that they acclimatize the forms. It is also true that we recognize that in a full-featured decennial census the possibility of conducting a 24-hour acclimatization of forms was not realistic or feasible.

And that is why in our proposal to the Bureau to deliver an enhanced system we proposed to develop, if you will, an internal humidity compensation capability in the scanner, that would basically allow the scanner to dynamically adjust as the form was passed through the scanner for changes in the size of the form. We were familiar with that requirement, sensitive to it and proposed to address it in our enhanced system.

Ms. **Fernandez.** Peter, it appears that probably the main reason why the Bureau has not moved to continue following up on the OMR improvements and modifications is the constraints of time. It seems that that is a perennial problem. As you know, on April 18 of last year we had the Census Bureau, the General Accounting Office, and Mr. Funk, inspector general of the Commerce Department, testify. And consistent with GAO's testimony today was Mr. Funk's testimony that unless there is a long term planning function at the Bureau, this is going to be a cyclical problem and that we can have the same expectations in the year 2000.

What is the Bureau planning to do to take it out of that cycle of planning for the immediate decennial, which brings to bare time pressures, time constraints, money constraints? When does the Bureau plan to move to planning for the next decennial, that does not have those time pressures and would allow you to explore modified versions, do demonstrations in various areas?

Mr. **Bounpane.** I think you have certainly put your finger on a very difficult problem, which is that there are limited resources in any agency, and I am not just talking about money, I am talking about people, as well, and when you get up against the crunch of doing the census it is very hard to think about tomorrow.

One way we are trying to solve that is through a current proposed reorganization which would set up some positions outside—not completely outside—but somewhat apart from the mainstream Census operations, and to put the development of the research and evaluation program for the year 2000 in this new arm. That should have some advantage— it will still be internal conflict, of course, for the people to work on it, but it will give an opportunity for some people to be thinking about the future and not be burdened with the day-to-day activities of making sure that 1990 works right. And I hope that change will help a little bit in this problem or we will be in the same position again.

Ms. **Fernandez.** If I recall, the inspector general also suggested that maybe the Census Bureau should look to outside experts to give them this technical advice. I think one of the common themes of both the testimony of the General Accounting Office and NCS is that as with all bureaucracies, we tend to hold on to what we have—it is risk-free and it has been demonstrated on a large scale decennial and that we don't want to shake the boat. There is also self-interest in terms of the manufacturing of the FACT 90 machine because it is manufactured in house and there may be some resistance there towards other automation systems that may be manufactured outside.

How will that be dealt with? I view that also as a particular problem that may not be solved by the reorganization that you just mentioned?

Mr. **Bounpane**. Yes. That, again, is a good point. Let me tell you what we do with regard to the American Statistical Association, and perhaps something parallel to that should be done with the set of private sector vendors.

In developing the research and evaluation program for the 1980 Census, we did ask our statistical colleagues to come in and make suggestions to us about things we should test in 1980 toward 1990. But it was in the limited arena of how best to take the census and statistical aspects of it.

Perhaps we need to do something parallel with the private sector, which says tell us what we should test in 1990 towards the year 2000 with regard to equipment available to us to take and process the census. The same kind of advisory approach with a panel like that might help in developing our year 2000 plans and I think that is something we will pursue.

Mr. **Franke**. Mr. Chairman, may I take off my NCS hat for just a moment to speak as a member of this industry. We have enjoyed our association with the Bureau of the Census. They have some challenges that we in private industry do not get to see very often, especially in the area that we are involved in.

I think that you would find that by going to the private sector, Peter, you would get a rousing response. We are out there, we are very interested, we have some good technology and we have some good people. I think it is a function of how that relationship is structured and where we are going to go with it that is going to dictate the degree with which the private sector is willing to commit to it. And I think if we can together—if you can set down some goals and invite us in to comment, to be involved in it—and I am certain I speak for most of the companies that are in this environment—we would welcome that opportunity.

Mr. **Garcia**. Mr. Dodaro.

Mr. **Dodaro**. Mr. Chairman, we would also support that. When Mr. Bounpane mentioned earlier, too, the strain on resources and constraints within the Department, I think it is only logical that a well-articulated planning strategy for where they want to go with the year 2000 ought to involve some active continuing advice from the private sector.

Mr. **Garcia**. If I may make a suggestion to the three organizations who are represented here today. I wholeheartedly agree, and this subcommittee would be only too happy to be the fourth part of that. In fact, there can be congressional input if some obstacles do arise and we are in a position to help. I think that that is the purpose of these hearings—

I think one of the other long term gripes I have with the Bureau of the Census is not any person at the Bureau, but because the politics is always there and there is a new director for each decennial. I would like to just take one director and keep him there forever. [Laughter.]

I mean that, because those decisions come from the top and whoever he or she may be—we shouldn't change the director constantly. What happens is the middle management of the census, which is there all the time become so accustomed to seeing directors coming and going that they question when is this person going to leave. I just don't think that speaks well for an agency, especially for one that disseminates information which is so absolutely vital for the health and the future of this country.

I think that one of the things that I am going to push for is to have a person at the head of the Bureau of the Census who is free of that political consideration, Democrat or Republican. They can just stay there, as a professional, and run the operation so that the people below them know that there is somebody there who will make a decision; and that that decisionmaker will be there in 2, or 3 years, or 5 years, or 10 years—however long they may decide that that is what they want to do. And I really feel strongly about that.

Having said that, I would like to, again, thank the three organizations who are present. I think this is what this subcommittee's aim is—to bring you together in honest, open and frank dialog. If we can work together without concern that we are looking to be critical of one another, or each other, and that we are really working in a good, solid framework I think we can have a good census in 1990 and we can have a Bureau of the Census that will be able to provide this country with quality data.

Thank you very much, I really appreciate your presentation. I know how busy all of you are. I have some questions that counsel will be submitting to the three of you, and I would appreciate it very much if you would look over the questions. As we have done in the past, if you can get those answers back to us, we can complete this record that we have established here today.

Thank you very much.

[Whereupon, at 12:15 p.m., the subcommittee recessed to reconvene at the call of the Chair.]

○